

対話構造を考慮した次発話予測手法の検討*

巖寺 俊哲 石崎 雅人 森元 暉†

3B-7

ATR音声翻訳通信研究所‡

e-mail: {iwadera,ishizaki,morimoto}@itl.atr.co.jp

1 はじめに

音声認識結果には、統語的、意味的に正しいがその文脈中では、不適切な候補が存在する場合がある。このような候補を排除し、認識精度を向上させるためには文脈情報が必要である。また文脈情報、特に次にどのような発話が起り易いかに関する情報を候補生成時に積極的に用いることにより処理の効率化を図ることができる。

従来より次発話を予測する手法として、大別すると2つのアプローチが試みられている。1つは、プラン[1,2]や話題遷移モデル[3]などのあらかじめ定義された知識を用いるアプローチである。もう1つは、統計的手法を用いるアプローチである。これは、発話タイプ等の連鎖により文脈を記述し、この連鎖を統計的にモデル化した次発話タイプを予測する手法である[4,5]。

これらの先行研究において、次の問題点が指摘されている。

1. 質問、要求等の聞き手に働きかける発話に対する予測力が弱い[3,4]
2. 問い返しを考慮したモデルを用いると予測精度が低下する[5]

これは、対話は単なる発話の連鎖ではなくある種の構造を持っているが、それが適切に考慮されていないために生じている問題であると考えられる。より高い精度で次発話を予測するためには、この構造を考慮する必要がある。我々は、この対話の構造を考慮した統計的な次発話予測手法の検討を行なっている。本稿では、次発話予測時における対話構造情報導入の効果に関する予備実験とその結果について報告する。

2 対話構造

実際に収集された対話データを観察すると、「一人の話者が話し掛け、別の話者がこれに応答する」といったパタンのくり返しを見ることができる。また、これらのパタンが幾重にも入れ子状になって繰り返されている。我々が提案している対話構造（インタラクション構造とよぶ）は、このようなパタンを表現するためのモデルである[6]。

このインタラクション構造は、ACT, MOVE, EXCHANGEの3階層から構成されている。ここで質問や要求によって話し掛けることを「開始」、応答することを「応答」、相手の発話に対するあいづちや補足、評価を「補足」とよぶと、対話中にあらわれるパタンは、次のようになる。

1. 「開始」+「応答」

2. 「開始」+「応答」+「補足」

このようなパタンをEXCHANGEとよぶ。これは、対話中の対応関係、たとえばある質問発話とそれに対応する応答発話の対応関係を表現している。MOVEは、上記のパタンにおいて「開始」や「応答」といった単位であり、同一話者による連続する1つ以上の次に述べるACTから構成される単位である。ACTは、ほぼ一文に対応する発話の最少単位であり、1つの表層発話行為を表す。

前述した対話中に現れるパタンが入れ子状に生じた場合、対応する発話同士、例えば質問と応答が隣接せず、離れてしまう。しかし、対話を単にセグメントに分割することによって対話の構造を捉えるのではなく、上記のように階層構造で捉えることで、対話中の表現としては隣接していない発話の対応関係を表現することができる。

前述した問題点1は、「開始」にあたる発話の予測力が弱くなる、すなわち1つの対話のパタンから別のパタンへ遷移する部分で予測力が低下する問題である。また、問題点2は、対話のパタンが入れ子状に生じたときに予測精度が低下する問題である。これらは、対話のパタンの状態およびその連鎖の状態を組み入れて発話の連鎖を記述することにより予測精度が向上すると考えられる。

3 統計的手法による次発話予測

統計的手法を用い、対話のモデルを考える場合、IFT[4]やdialogue act[5]のように発話を何らかの方法でモデル化し形式的に表現する必要がある。ここでは、あらかじめ求められているコードブックから得られるトークンによって表現する。コードブックは、以下のように作成される。まず、コーパス中の各発話をベクトル表示に変換する。すなわち、対話中の*i*番めの発話 U_i を下記のように*n*個の属性からなるベクトルで表現する。

$$\text{発話 } U_i = (f_i^1 \ f_i^2 \ \dots \ f_i^n)$$

これらの属性は、各発話の特徴つけるものである。たとえば、前述した対話のパタンのどこに属するか、すなわち「開始」、「応答」、「補足」のいずれに属するか、話者交替が起っているか、などによって記述する。次にコーパスから異なるベクトルの集合を求める。この集合の異なるベクトルに異なるトークンを与えることによりコードブックを作成する。

ここでは、前述したトークンの*n*-gramを用いて次発話を予測する。*n*個の発話からなる対話*D*は、トークン t_i を用いて次のように表現される。

$$D = t_1, t_2, \dots, t_n$$

この対話*D*の統計的モデルを $P(D)$ とすると、次のように*n*-gramの確率で近似できる。

$$P(D) = \prod_{i=1}^n P(t_i | t_{i-n+1}, \dots, t_{i-1})$$

*A Study of the Method of the Next Utterance Prediction Using A Dialogue Structure Information

†Toshiaki IWADERA, Masato ISHIZAKI and Tsuyoshi MORIMOTO

‡ATR Interpreting Telecommunications Research Labs.

n 番めの発話を予測することは、

$$t_n = \max_t P(t | t_{n-1}, t_{n-2}, \dots)$$

と近似できる。さらにここでは、削除補完法 [7] を用いて次式で近似する。

$$P(t_n | t_{n-1}, t_{n-2}) = q_1 f(t_n) + q_2 f(t_n | t_{n-1}) + q_3 f(t_n | t_{n-1}, t_{n-2})$$

ここで f は、コーパスから計算される相対頻度、また $\sum q_i = 1$ である。

4 実験

次発話予測時における対話構造の効果を検査するために、対話構造の使用の有無と予測精度の関係を実験により調査した。対話構造の有無の影響を検査するために対話中の各発話を表現するベクトル中の属性として対話構造に関する情報を含むものと含まないものの2種類を用意した。1つは、3つの属性、対話構造情報、対話ボタン中の位置、話者交替情報からなるベクトル、もう1つは、2つの属性、対話ボタン中の位置、話者交替情報からなるベクトルである。対話構造情報は、対話ボタンの埋め込み構造が生じているか否か、すなわち問い返しがおきているか否かを表す情報である。生じているか、いないの2値をとる。対話ボタン中の位置は、発話がボタン中のどこに属するかを表す情報である。すなわち、[開始]、[応答]、[補足]のどこに属しているかを示す。話者交替情報は、話者交替が起きているか否かを表す情報である。おきている、いないの2値をとる。

実験には、旅行に関する対話100対話を使用した。これらの対話を統計的学習と予測精度測定の対象にした。これらには、人手により対話中のどの発話がどのボタンを構成しているか、各発話は、ボタン中のどの位置に属するか、に関する情報が付与されている。また、話者交替情報が、自動付与されている。

次発話の対話ボタン中の位置の予測精度を実験により測定した。その実験結果を表1に示す。表中の値は、1位から3位までの平均累積精度を示している。この実験において、上述した100対話のうち、50対話を学習用データとし残りの50対話をテストデータとした。この

表 1: 実験結果

順位	学習データ		テストデータ	
	対話構造		対話構造	
	なし	あり	なし	あり
1	54.83%	57.34%	55.96%	56.02%
2	83.60%	85.92%	82.10%	85.00%
3	96.68%	96.73%	96.09%	97.15%

結果では、対話構造情報を導入することによりわずかに予測精度が向上している。

図1に[開始]に関する予測精度の測定結果を示す。この結果は、第2位までの累積予測精度を見ると約20

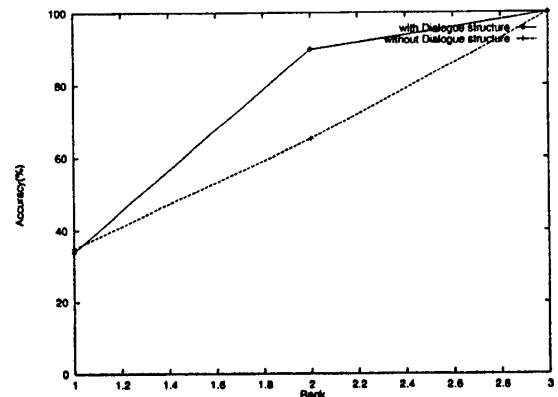


図 1: [開始] の予測精度

％向上している。この結果から、指摘されている問題点を解決するために対話構造に関する情報が寄与する可能性があることがわかる。

5 おわりに

本稿では、次発話予測時における対話構造情報導入の効果に関する予備実験とその結果について報告した。実験結果は、対話構造情報を用いることにより予測精度がわずかに向上することを示している。また、指摘されている問題点を解決するために対話構造に関する情報が寄与する可能性があることがわかる。この結果を詳細に分析し、対話の状態や発話をより適切に記述することによりさらに予測精度を向上させる手法を検討する。また、ここでは、3種類の次発話タイプを予測するだけであったが、今後は、より詳細な情報を予測する手法についても検討する。また、次発話を予測することが音声認識等ほどの程度寄与し得るか評価していく予定である。

参考文献

- [1] 山岡 孝行、飯田 仁：“階層型プラン認識モデルを利用した次発話予測手法—話し手の意図を表す表現についての音声認識結果曖昧性の解消”、電子情報通信学会論文誌、vol.J76-DII, No.6, 1993.
- [2] 山本 哲也、太田 義一、山下 洋一、溝口 理一郎：“音声理解システム SPURT-I のための対話管理機構:MASCOTS”、第4回人工知能学会全国大会、13-2, 1990.
- [3] 平松 敬史、吉田 英昭、野村 康雄、山下 洋一、溝口 理一郎：“音声対話理解のための話題知識の利用”、電子情報通信学会研究会、SP92-110、1992.
- [4] Nagata, T. and Morimoto, T.: “An Information-Theoretic Model of Discourse for Next Utterance Type Prediction”, 情報処理学会論文誌、Vol. 35, No.6, 1994.
- [5] Reithinger, N. and Maier, E.: “Utilizing Statistical Dialogue Act Processing in Verbmobil”, ACL-95, 1995.
- [6] 巖寺 俊哲、石崎 雅人、森元 暉：対話のインタラクション構造と話題の認識、情報処理学会自然言語処理研究会、NL-104-16, 1994.
- [7] Jelinek, F.: “Self-Organized Language Modeling for Speech Recognition”, In Waibel, A. and Lee, K.-F., editors, *Readings in Speech Recognition*, pp.450-506, 1990.