

韻律情報を用いた発話文タイプ付与支援システム

6H-2

青柳 達也 山本 幹雄 板橋 秀一

筑波大学

1 はじめに

近年、対話中の発話に対して、その発話の役割や話者の意図を表わす発話文タイプを付与し、音声認識器の性能向上や対話のモデル化に利用する試みが行われている[1]。しかし、膨大なデータに対して、人手でこれらの発話文タイプを付与するのは、非常に困難な作業である。そのため、発話文タイプを付与するシステムが必要となる。そこで、本稿では、そのようなシステムをつくる基礎技術として、発話文タイプを自動推定する手法を提案する。推定には、アルゴリズムC4.5[2]により生成される決定木を利用する。

C4.5で用いた情報は、形態素、音声情報（韻律・発話間のポーズ長）、及び話者交替といった表層的な情報である。推定の結果、テストデータで、発話内行為に対し93.5%、ムードに対しては94.5%の適合率を得た。

本稿では、まず推定手法を述べ、次に実験結果を示し、最後に考察を述べる。

2 発話文タイプの推定手法

2.1 アルゴリズム

C4.5とは、訓練データから、自動的に決定木を生成するアルゴリズムの一つである。本稿では、C4.5を用いて、訓練データから発話文タイプ推定のための決定木を生成し、テストデータについては、その決定木を利用して発話文タイプを推定する。

2.2 発話文タイプ

発話文タイプとして、次の2種類を細分類したものを設定した。詳細は表1、表2を参照。

- ・発話内行為：前後の文脈から推測される
- ・ムード：文法から推測される

2.3 表層情報

以下の4つの表層情報を用いた。

- (1) 文末の形態素列
- (2) 韻律情報（音声の基本周波数 F_0 、音声power）

韻律は、話者毎に正規化し、折線近似[3]を行った。そして、最後の一本の折線の傾き、開始点値、及び終了点値を用いた。

A semi-auto IFT tagging system using prosodic information.

Tatsuya AOYAGI, Mikio YAMAMOTO,

Shuichi ITAHASHI

University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki 305, JAPAN

(3) ポーズ長

推定したい発話単位と、その直前の発話単位との間のポーズ長を用いた。なお、相手の発話の中で発話が行われる発話内発話に関しては、ポーズ長は「その他」とした。

(4) 話者交替

推定したい発話単位と、その直前の発話単位との間に話者交替が「ある」/「ない」という情報を用いた。なお、ここでも発話内発話に関しては「その他」とした。

表1 発話内行為

発話内行為	頻度	割合(%)
情報伝達	400	35.7
真偽情報要求	153	13.6
未知情報要求	87	7.8
儀礼	36	3.2
行為要求	19	1.7
行為提案	12	1.1
その他	413	36.9
合計	1120	100

表2 ムード

ムード	頻度	割合(%)
確言	243	21.7
真偽疑問	113	10.0
依頼	70	6.3
疑問語疑問	58	5.2
願望	25	2.2
命令	11	1.0
否定	8	0.7
選択疑問	2	0.2
意志	1	0.1
概言	1	0.1
その他	588	52.5
合計	1120	100

3 実験

3.1 実験データ

文部省重点領域研究「音声対話」の対話音声コーパス Vol.1中の8対話、及びRWC音声対話コーパス中の2対話の計10対話を用いた。対話データの概要を以下に示す。

- ・全発話単位数：1120
- ・一対話あたりの平均発話単位数：112
- ・全形態素数：8826
- ・一発話単位の平均形態素数：7.88

上記の全発話単位に発話文タイプを付与した。その頻度と割合を表1~2に示す。付与作業は筆者が行った。

3.2 実験方法

実験は、10対話のうち9対話を訓練データ、残りの1対話をテストデータとする、クロスバリデーションにより行った。

3.3 パラメータを決定する実験

次の四つのモデルから、表層情報についてのパラメータを決定した。パラメータの値を変化させ、各々の値でクロスバリデーションを行い、その中でテストデータに対して最良の結果となる値を採用した。

- ・モデル1：推定対象の発話単位の形態素のみ使用
- ・モデル2：推定対象の発話単位と直前の発話単位の形態素を使用
- ・モデル3：Foのみ使用
- ・モデル4：powerのみ使用

パラメータは、文末からの形態素数は、発話内行為では4、ムードでは5にした(モデル1・2)。また、Foの折線近似の閾値を74(モデル3)、powerの折線近似の閾値を120(モデル4)にした。なお、この閾値とは、元の波形データと、近似した折線との間の最大距離である。

3.4 推定実験

上記実験で設定されたパラメータを用い、以下の5つのモデルで推定実験を行った。

- ・モデル5：韻律情報のみ使用
- ・モデル6：音声情報のみ使用
- ・モデル7：音声情報以外を使用
- ・モデル8：全ての情報を使用
- ・モデル9：現在の発話単位の形態素と音声情報を使用

4 実験結果

図1に各モデルの適合率を示す。なお、モデル1~4に関しては、最良の結果のみ示してある。全ての情報を用いたモデル8と、テストデータに対して最も良い適合率が得られたモデル9の結果を、表3と表4に示す。

表3 モデル8での実験結果

	訓練データ	テストデータ
発話内行為	95.8	89.7
ムード	95.8	93.4

表4 モデル9での実験結果

	訓練データ	テストデータ
発話内行為	95.3	93.5
ムード	95.4	94.5

5 考察

訓練データでは、全ての情報を用いるモデル8が、発話内行為、ムード共に最も良い適合率となっている。一方、テストデータでは、モデル9の方が、モデル8よりも発話内行為、ムード共に良い結果を示した。この理由としては、訓練データが少ないため、前の発話単位の形態素の情報がノイズになってしまっていることが考えられる。訓練データでは、モデル8の方が若干ながら良い結果を示しているため、訓練データを増やせば、テストデータに対してもモデル8の方が良い結果を示すと考える。

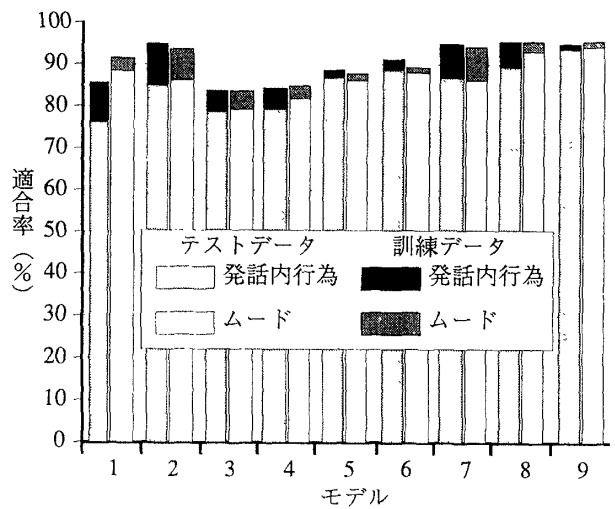


図1 各モデルでの実験結果

6 おわりに

韻律や形態素などの発話の表層情報を用いて発話文タイプをC4.5により自動推定する手法を提案した。C4.5で生成される決定木を用いることによって、テストデータで、発話内行為に対し93.5%、ムードに対しては94.5%の適合率を得た。

今後の課題としては、推定したい発話単位の直後の発話単位の情報を用いるモデルの作成や、数量化理論第II類などの他の手法での推定実験、及び現在制作中の支援システムの完成と評価が挙げられる。

参考文献

[1] 石崎雅人, 小磯花絵, 「対話研究の新しい流れ」, 人工知能学会研究資料, SIG-SLUD-9503-1, (1996)
 [2] J.R.Quinlan, 古川康一訳, 「AIによるデータ解析」, 株式会社トッパン, (1995)
 [3] 板橋秀一, 「基本周波数パタンの折線による記述-無声音を含む場合-」, 音講論, 3-1-1, pp.479-480, (1977)