1—8—26

対話音声における非語彙的情報の統計的モデル化の検討

◎田中俊光(奈良先端大) △柏岡秀紀(奈良先端大/ATR) ニック キャンベル (奈良先端大/ATR/CREST)

1. はじめに

現在までの音声処理では主に音声情報と言語情報との関連を見るものが主流である。また数年ほど前から韻律情報の研究が盛んになってきた。しかし、対話処理において理解や笑いという話者心理に近い高度な情報を扱うにはいたっていない。理解や笑いなどの発話の種類を統計的なモデルで扱うことができれば、ユーザへの負担がない入力を実現できる。本稿では、統計的なモデルとして決定木を用い、韻律情報と発話の種類との関係について議論する。データには、CREST-ESP で収集された電話対話データを利用した。このデータには、発声の種類がラベルで付与されている。韻律情報、前発話の種類を説明変数として利用した決定木による発話の種類の自動ラベリングの実験について述べる。

2. 自動ラベリング実験

本節では、決定木を用いた発話の種類の自動ラベリング実験について述べる。

2.1 使用データ

本実験では CREST-ESP において作成されたデータを使用した。このデータは、電話対話の音声であり、書き起こし及び発声種類のラベルが発話毎に付与されている。ラベルは、聞いている指標としてのあいづち(以下あいづち)、フィラー、笑いなど 34 種類以上ある。同一の発話に複数の発声種類ラベルが付与される場合には併記されている。ラベル付与は 3 人のラベラーによる。特に短い発話時間長の場合は聞いているという意思表示のあいづちが非常に多い特徴がある。

実験には、片話者のデータを使用し、女性話者 1 名について相手話者が男性女性各 10 対話のデータを用いた。ラベリング対象とした話者の発話数は女性から女性への 2545 発話と女性から男性への 3503 発話の計 6048 発話である。次節で示す使用ラベルの分布を考慮すると、発話時間長が 0.7s 以上(3028 発話)の場合に、各ラベルがほぼ均等に出現する（ただし、気付きは、非常に短い発話にしか付与されていないため、均等にすることができない）ことから、全体での分析と、0.7s 以上のデータでの分析を行った。

2.2 使用ラベル

ラベルの中で、頻度が高く、ラベリングが有意義なラベルと思われるラベルに限定して自動ラベリングを行った。本稿で扱ったラベルは、あいづち、フィラー、笑い、理解、気付き、質問の 6 種類である。ラベル選定についてはあらかじめ対話全体を聞き人手によって付与されたラベルをさらに当該発話のみの聴覚的特徴の似ているラベル、あるいは意味的に類似していると判断できるラベルについて再ラベルを行い、そのラベルを使用した。

2.3 決定木

人手によるためラベルの揺れを考慮し、比較的ノイズに強い手法として決定木を用いた。

説明変数には韻律情報として、F_0、パワー、時間長、及び発話速度を用いた。F_0 については平均、最大値、最小値、最大最小の差、発話時間で正規化した最大と最小のそれぞれの位置、最大値から最小値への傾きの 7 つ、パワーについても同様の 7 つの値を使用した。またラベリングの際に大きな影響を及ぼす文脈への依存度を考慮し、予測対象の前の発話ラベルも説明変数へ追加した。

2.4 実験手順

対人の性別による声の使い分けの存在が一般に知られていることから以下の二つの実験を行った。

・対人の性別を考慮しないラベリング実験
・対人の性別を考慮したラベリング実験

3. 結果

3.1 対人の性別を考慮しないラベリング

対象データ全体で決定木学習した場合と、発話時間長が 0.7s 以下と 0.7s 以上に分けて決定木学習した場合の、自動ラベリング実験を行った。また、2.3 に述べた予測対象の前発話のラベルを説明変数に追加することによる全体の精度上昇は 1-2%程度にとどまった。表 1 に、0.7 秒以下、0.7 秒以上、全体について、それぞれ closed および 10 分割の cross-validation を行った際の前発話のラベル追加後の自動ラベリングの予測精度を示す。またその下欄に最頻度のラベルの出現割合と発話数を示す。

表 2 に、ラベルの分布を考慮し、0.7 秒以上の場合および全体での場合について、ラベル毎の精度について示す。

"A study of statistical modeling of non verbal information in conversational speech" by T.Tanaka(NAIST), H.Kashioka(NAIST,ATR) and N.Campbell(NAIST,ATR, CREST)

表1：決定木によるラベリング精度

	0.7s以下	0.7以上	全体
Closed	86.7%	77.1%	82.4%
cross-validation	80.0%	60.5%	70.1%
最頻度ラベル	69.4%	22.4%	45.9%
発話数	3020	3028	6048

表2：ラベルごとのラベリング精度

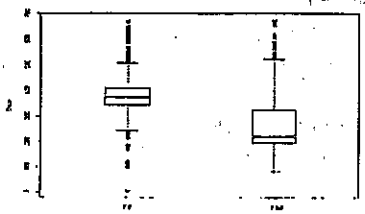
ラベル	全体	0.7s以上	ラベル	全体	0.7s以上
あいづち	80.8%	65.4%	フィラー	59.2%	62.2%
笑い	71.8%	78.2%	理解	58.8%	62.2%
気付き	53.9%	-	質問	64.1%	68.9%

笑いに関して予測精度が高いことが分かる。また気付きが 0.7s 以上の精度が無いのは該当箇所にラベルが存在しないためである。

3.2 対人の性別を考慮したラベリング

はじめに、対人の性別による日常会話での声の使い分けについて、決定木に利用した説明変数に対する発話の分析を行った。おおまかな違いを見るため、箱ひげ図(図 1：F_0 最大値について)を示す。他の多くの説明変数でも対人の性別が異なる場合に、中央値に有意差がみられた。特に F_0 については顕著な違いがあることを確認した。

図1．あいづちの F_0 最大値の箱ひげ図
(左：女⇒女、右：女⇒男)



次に、対人の性別を考慮した発話の自動ラベリング実験を行った。対人の性別ごとに決定木を作成し、テストセットに、学習データと同じ条件のデータを使用した場合と、異なる場合の結果を表 3 に示す。女性(F)から男性(M)を FM とし、他の条件も同様に表記した。

表3：性別の異なるテストセットによる比較精度

Test set	FF	FM
FF	72.1%	66.4%
FM	66.7%	66.2%

これらの結果から同一ラベルにおいても女性から女性の発話の方が比較的落ち着いた発話である。逆に男性へ話す場合、特に決定木の上部に高頻度で現れる F_0 最大値などが幅広い範囲をもつ発話である。そのため女性同士より精度も落ちる。

4．考察

全体 45.9%の最頻度ラベルが存在する中でのラベリング精度 70.1%は良い結果とは言いにくい。笑いのラベルについては非語彙的な韻律情報から 7 割の比較的高い予測精度を持っている。対人の性別を考慮した実験から、女性話者から男性の方が幅広い発話であることが伺える。そのため自動ラベリングの精度向上には相手の性別あるいは対人関係を考慮したさらに詳細な分析を行うことが有効であると考える。

5．まとめ

女性 1 話者において自然対話における発声種類を示すラベルを統計的な手法に基づき自動ラベリングした実験の結果を示した。ラベルごとの評価から、笑いのラベリング精度は高いことがいえる。また対人の差異による影響を見るためのラベリング実験を行い、精度に差異のあることを確認した。本研究は自動ラベリングを行う統計的モデルを詳細化、高度化するための研究の第一段階である。

6．検討課題

女性話者一名の分析であるため、今後、個人性の考慮が必要である。そこで、複数の話者について実験、検討を計画している。また女性話者から男性の方が韻律の幅が広い発話であった。これが異性間の場合の共通の傾向であるのか、あるいは男女では同性異性への声の使い方に違いがあるのか調査を行っている段階である。直前自発話との間のポーズ長、対人の違いを見る際には F_0 と相関のない NAQ といわれる正規化した声の硬さを表す指標を導入することも検討すべき課題と考える。

謝辞

本稿作成にあたり、データ提供をして戴いた CREST-ESP にも感謝の意を表したい。特にラベラーの木村さんにはラベルについての解説を丁寧にして頂き感謝いたします。また所属研究室の皆様には貴重なアドバイスをいただき参考になりました。ここに感謝申し上げます。

参考文献

[1] 芳村、キャンベル：JST/CREST 発話様式プロジェクトの電話対話データベース、人工知能学会全国大会論文集 3C5-11(2002)

[2] 藤尾、キャンベル、樋口：韻律を用いたテキスト非限定型発話アクト識別方法，音講論秋季，1-4-14(1996)

[3] 田中脇本：多変量統計解析法，現代数学社，(1983)

[4] 前川：パラ言語情報の知覚(p13-16)，「韻律に着目した音声情報処理の高度化」研究成果報告書，(2001)