

発話間単語共起関係に基づく補正信頼度を用いた音声認識

松本 宗也，谷口徹，白井克彦（早稲田大学理工学部），

樽松明（早稲田大学理工学総合研究センター）

1. 序 論

携帯端末をはじめメディアが年々高度化する中で、音声技術の応用範囲も広がってきている。特にハンドレス・ユーザフレンドリ・バリアフリーに関して注目されるようになってきており、これらに対し音声認識を用いることは有効な手段の一つであるため、より実用的な音声認識システムが求められてくる。音声認識システムは円滑に目的を達成できることが重要であり、必要に応じて確認したり聞きなおしたりするといった柔軟な対応を行う必要があるため、特に対話システムを構成する音声認識システムは、より重要な役割を担ってくる。このとき、音声認識結果に対しどの程度信頼してよいかを表す信頼度尺度を用いることで、後段の音声認識処理に対してその値を考慮した処理を行う事ができるため [1]、対話システムを実現するにあたりそういった尺度を用いることは有効に働く。

従来の音声認識では音響的特徴（HMM）や言語的特徴（n-gram）に関しては多く研究されておりその効果も十分に期待できるが、話している内容を考慮したようなモデルについては未だに効果的な手法がない状態である。音声認識の信頼度と対話履歴を利用してユーザ発話意図の推定を行う研究 [2] では、認識結果に対しカテゴリ識別を行う事で内容を特定し、頑健な音声認識を行う事を目標としている。この場合、信頼度はシステムを構築する上で重要な役割を果たしているが、対話をカテゴリに分類するため、柔軟な対話をさせたり、自然な対話に対して期待する精度を得ることは難しい。本研究では柔軟な対話システムの実現を目指し、よりよい信頼度尺度を提供するための手法を提案する。信頼度を考慮した処理を行う場合、その精度が対話システムに対し大きく影響を与えるため高い精度が求められる。より精度の高い信頼度を得るためには話している内容を考慮する必要があるため、本研究では発話内容を考慮することでより精度の高い信頼度を提供する方法を考えた。

2. 提案手法

2.1 提案手法の概要

一般的に対話は交互に行われる事が多いため、発話内容に対する応答がある程度推測できる。そこで本研究では、対話の内容を発話間の単語共起関係により考慮し、共起情報を予め対話コーパスから統計的に学習しておく

ことで音声認識に利用することを提案する。具体的には、システム側の発話文に対する応答文の認識において、発話文に応じた認識文の共起情報を単語毎に動的に取得する。この共起情報を利用して音声認識システムから得られた単語の信頼度を補正し、補正信頼度として提供する。提案手法のイメージを図 1 に示す。使用する音声認識システムは、大語彙音声認識エンジンである Julius[3] を使用し、Julius から得られる単語毎の信頼度を用いる。

2.2 コーパス学習の方法と利用

共起データの取得方法について説明する。学習に使用したコーパスは、スケジュール、特に出張に関する話題について二人が対話したものである。二人の話者を A、B としたとき、A の先行発話中の各単語に対し B の応答文中に出現した各単語をカウントする。このとき、最長一致法により形態素解析を行った。

例) A: 明日は晴れ?

B: たぶん 雨

上の例で『明日』、『は』、『晴れ』という先行発話の各単語に対して、応答文の単語である『たぶん』や『雨』に対してそれぞれカウントを行う。得られた共起情報に基づき、発話文と認識単語の関係を表すスコアを算出する。

認識した応答文中の単語 word に対するスコア $S(\text{word})$ を求める式を以下に示す。ただし、 w_k は質問文中の k 番目の単語であり ($1 \leq k \leq N$)、単語毎の統計量のばらつきの影響を抑えるため、単語 w_k に対し各単語の条件付確率の平均値 P_{ave} で正規化したものを使用した (式 2)。M は質問文中の単語 w_k における登録単語数 (学習時に一回以上出現したことのある単語数) である。

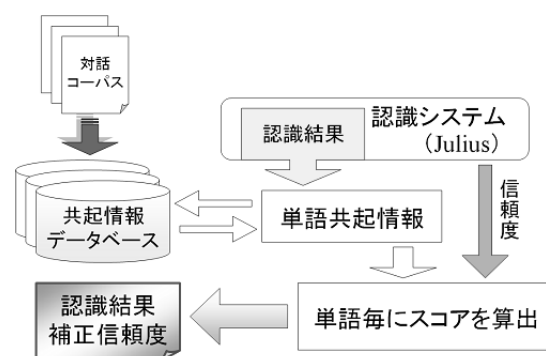


図 1 提案手法のイメージ

表 1 Julius 実験条件

音響・言語モデル	詳細
高精度成人モデル	ATR 多数話者音声データベース より学習 PTMtriphone モデル 3000 状態 64 混合
60K 単語 3-gram	Web から収集したテキスト 形態素 : chasen2.3.3+ipadic2.6.3
学習用対話コーパス	VERBMOBIL-2 (Data Collection) 出張に関する話題 (約 170KByte)
学習用一般コーパス	青空文庫 5 冊 (約 150KByte) Web ニュース記事 5 件 (約 30KByte)

表 2 実験に用いた認識用データ

(A) コーパス依存対話	20 文 (計 160 発話) 学習した対話コーパス (出張) の一部を含んだ対話応答文
(B) 一般スケジュール対話	10 文 (計 80 発話) スケジュール全般の対話応答文
(C) 一般対話	10 文 (計 80 発話) 全く自由な話題の対話応答文

ジュールに関する話題を除いた全く自由な対話である。

3.3 評価方法

対話では誤認識したままでは問題があるような場面も多く、正確な認識が求められることもある。そこで、評価基準として認識率に加え、正確度を用いた。認識率は発話文に対する正解単語数の割合を表し、正確度は受理した結果の中での正解した単語の割合を表している。認識率は単にどの程度の割合で正しく認識できているかを表すが、どの程度誤認識が含まれているかといった情報は、認識率からは分からないため、正確度によって受理した認識結果の中でどの程度正確に認識しているかの割合をみる。

認識単語集合を W とし、 $U(W)$ を認識文全体の単語数、 $C(W)$ を認識文中において正解した単語数とすると認識率 $P(W)$ は次のように求まる。

$$P(W) = \frac{C(W)}{U(W)} \quad (6)$$

更に棄却単語数を $D(W)$ とし、正確度 $Q(W)$ を次のように定義する。

$$Q(W) = \frac{C(W)}{U(W) - D(W)} \quad (7)$$

得られた信頼度に対してその閾値を変化させ、0.00 ~ 0.20 の間の 10 点において閾値未満の単語を棄却し、その認識率と正確度の変化をみることで評価を行う。

3.4 Julius 信頼度の評価

音声認識実験により Julius 信頼度の評価を行った (図

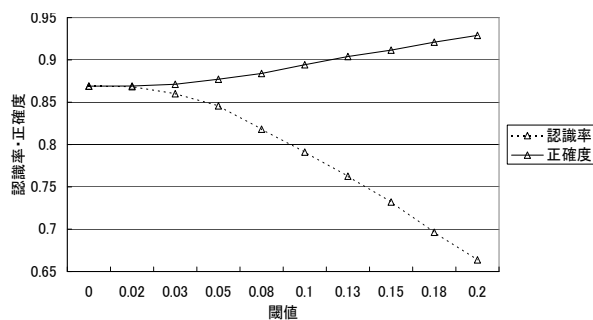


図 3 Julius 信頼度における閾値と認識率・正確度の関係

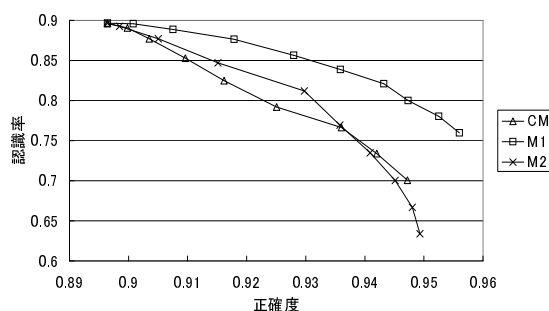


図 4 信頼度 CM と補正信頼度 M1 (対話コーパス学習), M2 (一般コーパス学習) の比較 (実験データ (A))

3). 実験データは A ~ C (表 2) の 40 文である。グラフより、閾値を上げていくと認識率は低下するが正確度は上昇する傾向がみられた。このことから高い信頼度のものは正確に認識できている事が分かるが、閾値棄却において認識率が大きく低下していることから正しく認識できている単語も低い信頼度で与えられているものが多いことが分かる。閾値棄却を行う場合、正確な認識は可能になるが認識率そのものは大きく下がる。

3.5 Julius 信頼度と補正信頼度の比較

Julius の信頼度と補正信頼度を比較するため、Julius 信頼度 CM, 対話コーパス学習による補正信頼度 M1, 一般コーパス学習による補正信頼度 M2 に対して評価を行った。この時、3 種類のデータ (表 2) について別々に実験を行った。それぞれのデータ (A) ~ (C) における実験結果を図 4 ~ 6 に示す。ここで、閾値に対する認識率・正確度のグラフでは認識率が低下し正確度が上昇するといったケースの比較を見やすくするため、グラフは認識率と正確度を縦軸、横軸にとったものとした。グラフを見ると、図 4, 5 において対話コーパス学習を行った補正信頼度のものが認識率・正確度共に上昇し、良い結果が得られている。しかし、学習したコーパスとは全く関係のない話題では (図 6) 補正信頼度の方が一部悪化していることから、共起学習を行うコーパスの話題と実際に認識を行う時の話題は関係があることが分かった。

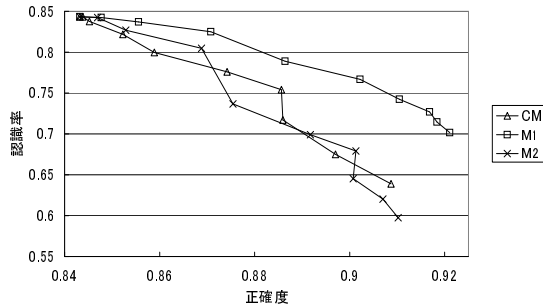


図5 信頼度 CM と補正信頼度 M1 (対話コーパス学習), M2 (一般コーパス学習) の比較 (実験データ (B))

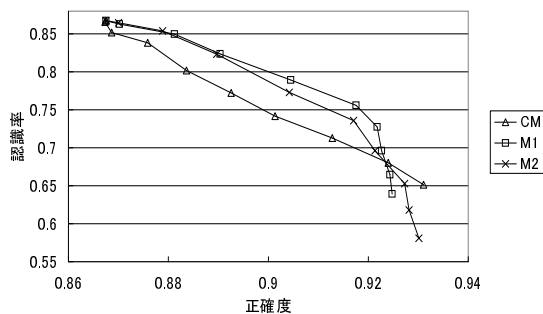


図6 信頼度 CM と補正信頼度 M1 (対話コーパス学習), M2 (一般コーパス学習) の比較 (実験データ (C))

3.6 通常認識と重要単語認識の比較

対話において相手の意図を理解するためにはある特定の単語を認識するだけでよく、相槌や一部の助詞に関して認識する必要はない。そこで、認識結果のうち自立語を中心とする意味を理解する上で必要と考えた単語を主観により重要単語とし、重要単語だけの認識率・正確度を算出して通常認識との比較実験を行った。実験データは (A) を用いた (表 2)。図 7 は対話コーパス学習による補正信頼度での通常認識と重要単語認識の結果であるが、閾値棄却を行わない状態では重要単語認識の方が認識率が低い結果となった。しかし、棄却前で認識率・正確度が 3.1% 低かった状態から、閾値 0.20 での棄却後では認識率・正確度は 0.7% 低いだけに留まり、棄却による評価は高い結果となった。Julius 信頼度においても同様の実験を行った結果 (図 8)、一般的に重要単語の方がその信頼度の精度が高く、補正信頼度を使用した方がより高い精度が期待できる結果となった。

4. まとめ

本研究では、柔軟な対話システムの実現を目指し、より精度の高い信頼度尺度を提供するための手法として、発話間における共起情報の利用の有効性を調べた。実験の結果、近い話題において発話間の単語共起情報は有効に働いたが、異なる話題に関してはその効果は期待できないことが分かった。しかし、システム主体の対話にお

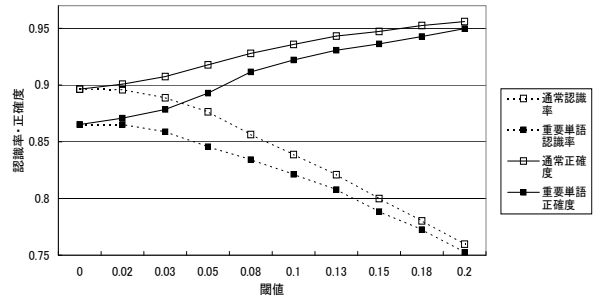


図7 通常認識と重要単語認識の比較

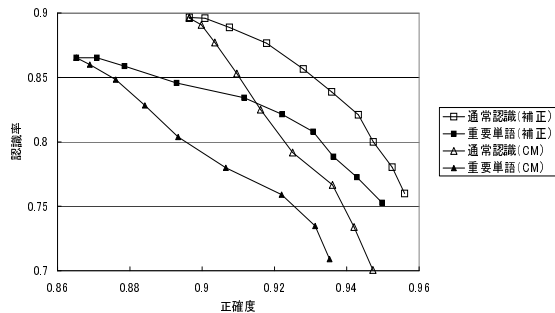


図8 補正信頼度・Julius 信頼度における通常認識と重要単語認識の比較

いて話題を特定することは容易であるため、予め複数の話題に対して学習しておくような応用も可能であろう。重要単語に着目した実験では、通常認識に比べ信頼度の精度が高く、また共起情報の効果も期待できることが分かった。対話システムの音声認識では特定の単語に着目すればよいので、今後はある話題に関するコーパスとそれとは異なる話題に関するコーパスで学習した共起情報を比較することにより、重要単語の抽出ができないかを検討していきたい。

謝辞 本研究の一部は、早稲田大学理工学総合研究センターの研究課題「自発的コミュニケーション機構を有するマルチモーダルヒューマンインタフェースの研究」、平成 17 年度科学研究費基盤研究 (B) 課題番号 17300066 「対話状況に応じた自発的コミュニケーション機構の研究」によるものである。

文献

- [1] 駒谷和範, 河原達也, “音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理”, 情報処理学会論文誌, Vol.43, No.10, pp.3078-3086, Oct. 2002.
- [2] 翠輝久, 河原達也, “頑健な意味理解のための音声認識信頼度と対話履歴を利用した発話意図推定手法”, 情報処理学会研究報告, 2005-SLP-55, pp.77-82, Feb. 2005.
- [3] 連続音声認識ソフトウェア Julius, <http://julius.sourceforge.jp/>
- [4] 李晃伸, 河原達也, 鹿野清弘, “信頼度基準デコーディングを用いた高効率な単語グラフ生成法”, 情報処理学会研究報告, 2005-SLP-55, pp.71-76, Feb. 2005.
- [5] 李晃伸, 河原達也, 鹿野清弘, “2パス探索アルゴリズムにおける高速な単語事後確率に基づく信頼度算出法”, 電子情報通信学会, NLC2003-97, SP2003-160, pp.35-40, Dec. 2003.