

決定リストを用いた局所係り受け解析と音声合成における ポーズ制御への応用

梅村 祥之 原田 義久 清水 司 杉本 軍司

豊田中央研究所

1 はじめに

近年、カーナビゲーションシステムを初めとする種々の情報機器が自動車に搭載され、交通情報、電子メール、ニュース記事等の情報提供サービスが始まりつつある。自動車では、音声による情報提示が望ましく、文章データを入力して音声波形に変換するテキスト音声合成技術の重要性が増しているが、品質面でまだ問題がある。中でも韻律（ポーズ、イントネーション等）が良くないと、不自然で棒読みな感じを与え、悪くすると意味を取り違えることにもなる。本論文は、入力文から、ポーズ挿入位置を判定する技術において中心的な役割を果たす係り受け解析法および解析結果に基づくポーズ挿入位置判定法に関するものである。

文から係り受け構造を求めるための、係り受け解析では、韻律制御用途には、局所係り受け解析で十分である。言語処理の基本技術である係り受け解析が盛んに研究され、機械学習の枠組みの中では、ME法による係り受け解析手法[1]が、最も高精度な手法と考えられている。しかし、学習によって得られた統計モデルを蓄えた解析辞書の容量を、設計の現場において削減することによりメモリ量と計算速度を調整するという事は容易ではない。また、素性を削減して統計モデルを再構築するには、学習に膨大な計算時間を必要とする[1]。そのため、車載情報機器や携帯情報端末など、小型化、低価格化に厳しい要求があり、開発サイクルも短い設計現場に不向きである。

そこで、ME法と同等の精度で、かつ、メモリ容量と実行速度の調整が容易で開発現場に受け入れられやすい、という特徴を持つ係り受け解析手法を開発するため、メモリ容量と実行速度に関して容易に設定変更ができ、アルゴリズムがシンプルで移植・保守の容易な決定リスト [2]を採用することにした。

係り受け解析結果に基づくポーズ挿入位置判定では、文の構文的な構造とポーズ、イントネーションとの関係に関する研究がなされ、構文構造に基づいてポーズ挿入位置を決定する研究がなされている

[3]。その結果、近傍文節間の係り受け関係がポーズ挿入位置の決定に重要であることがわかってきている。従来研究の中で、ポーズ挿入位置設定規則検討のための実験を最も大規模に行っているのは文献 [4]の研究である。この研究では、アナウンサ 10 名によって発声させた ATR 音声データベースの 503 文のポーズ長を分析して、ポーズ挿入規則を作成し、それを基にポーズ制御した合成音声 100 文と自然音声のポーズ長をそのまま使ってポーズ制御した合成音声 100 文を 10 名の被験者に提示してポーズ挿入規則の評価を行い、自然音声のポーズと同等なポーズ挿入規則が作成されたと報告されている。

本論文では、これらの研究から明らかになった、係り受け距離と句読点に基づくポーズ挿入規則をベースに作成した合成音声を用いて聴取実験を行い、悪い評価となった文を分析することによって、さらに追加すべき規則がないかどうか検討する。

2 決定リストを用いた局所解析

2.1 手法

本手法は、決定リストを用いて、文節間係り受け関係を判定している。特に文献[2]で述べられた手法をベースにして決定リストを構築する。係り受けを決定リストによって決定する方法として、従属節係り受けに決定リストを用いた解析[5]が提案されている。同文献に示された従属節係り受けの例を図 1 に示す。

我々の手法は、韻律制御において必要な局所係り受け解析に決定リストを適用したものである。局所係り受けのうち最も狭い範囲の解析としては、直後の文節に係るか、係り受け距離（文節数で距離を定義）2 以上に係るかを判定するタスクである（図 2）。

決定リストの素性の構成に関し、従来用いられているものを参考に、基本的構成（素性 1）と複合的構成（素性 2）の 2 タイプを定める。

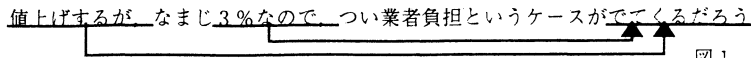


図1 従属節係り受け

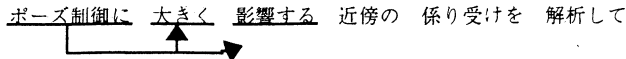


図2 近傍係り受け

素性1:

係り元文節に付属語がある場合、付属語の表記を素性とする。係り元文節に付属語がない場合、主辞の品詞、詳細品詞、品詞と活用形の組合わせ、詳細品詞と活用形の組合わせを素性とする。また、係り先文節の主辞の品詞、詳細品詞、品詞と活用形の組合わせ、詳細品詞と活用形の組合わせを素性とする。

ここで、主辞は、最後の自立語として定義する。品詞と詳細品詞は、Jumanの定義に従う。

素性2:

係り先文節に読点を含むかどうかの情報を素性とする。係り元文節に読点を含むかどうかの情報を素性とする。係り先文節の主辞が活用しないとき、係り先文節の主辞の品詞と次の文節の主辞の品詞の組合わせ、係り先文節の主辞の詳細品詞と次の文節の主辞の品詞の組合わせ、係り先文節の主辞の品詞と係り先文節に助詞を含めば助詞の表記と次の文節の主辞の品詞の組合わせ、係り先文節の主辞の詳細品詞と係り先文節に助詞を含めば助詞の表記と次の文節の主辞の品詞の組合わせを素性とする。係り先文節の主辞が活用するとき、係り先文節の主辞の品詞と活用形と次の文節の主辞の品詞の組合わせ、係り先文節の主辞の詳細品詞と活用形と次の文節の主辞の品詞の組合わせを素性とする。係り先の主辞の品詞が判定詞か接尾辞の場合、それらの前の形態素の品詞と活用形を素性とする。

／ ポーズ挿入の正解箇所の数

適合率 = ポーズ挿入の正解箇所挿入した数

／ ポーズを挿入した数

$F \text{ 値} = (1 / ((1 / \text{再現率} + 1 / \text{適合率}) / 2))$

ここで、「係り受け距離1ならポーズを挿入せず、2以上なら挿入する」という操作をポーズ挿入の正解と見る。F値とともに、係り受け解析において通常定義される正解率も算出する

- 局所係り受け解析を評価するにあたり、形態素解析と文節まとめ上げは100%を与える

(1) 人手による規則でのポーズ制御のF値

文献[1]で示された人手による簡単な規則による係り受け解析では、5,000文に対するF値は75.35%であった。

(2) 開発手法によるポーズ制御のF値

前述の素性1、素性2の決定リストを基にポーズ挿入位置の判定を行ったときのF値を図3に示す。

素性1による結果:

F値 = 88.18%, 正解率 = 92.22%

素性1と素性2を決定リストに統合した結果:

F値 = 90.04%, 正解率 = 93.33%

(3) 学習量によるF値の変化

学習の文数を1,000文から1,000文ずつ1万文まで増加させて正解率とF値を求め、図4に示す。

2.2 実験と結果

解析の性能評価を次の評価方法で行う。

- 京大コーパスの5,000文で評価する
- 「係り受け距離1ならポーズを挿入せず、2以上なら挿入する」という基準を正解とみなす
- 次の定義のF値により評価する

再現率 = ポーズ挿入の正解箇所挿入した数

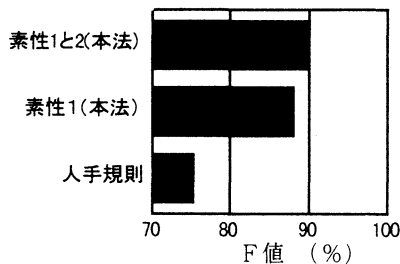


図3 各方式の精度 (F値)

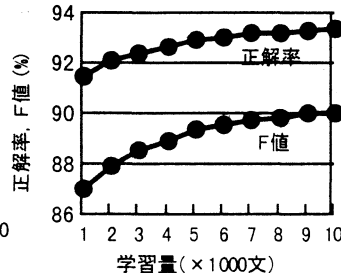


図4 学習曲線

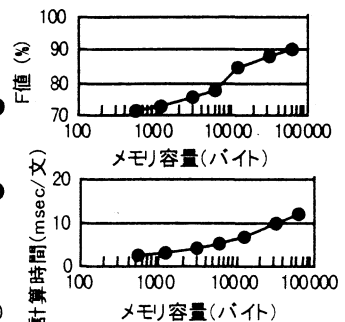


図5 メモリ容量と精度 (F値), 計算時間の関係

2.3 辞書容量の削減

統計的手法で高い正解率を得るためには、一般に辞書に多くのメモリを要し、計算時間もかかる。それに対して、本法は、メモリ容量と計算速度を、目的とするシステムのリソースの状況に合わせて調整することが可能である。以下にその方法を述べる。

辞書データ、すなわち、決定リストは、その原理上、判定のための証拠を重要な順に並べている。よって、メモリ容量を半減したければ、決定リストの上位から半分を残してそれ以下を削除すればよい。また、計算速度に関しても、決定リストのサイズを削減してゆくと、リストの上位から順に条件判定をする方式のため、単調に速度が向上して行く。

メモリ容量を減らして、計算速度を上げれば、その代償として正解率は低下する。以下に、決定リストのサイズを変化させて、正解率と判定時間を測定した結果を図5に示す。実験条件は次の通りである。

- 学習量は1万文固定
- 計算時間の算出は、実測によった。プログラミング言語はLISP、CPUはPentiumIII 450MHz。形態素、文節、決定リストがメモリに読み込まれている前提で、各文節から直後の文節への係り受けを判定し、ポーズ判定する時間を測定
- メモリ容量を算出するにあたり、決定リスト1行に現れる素性を表現するに必要な情報量から、1行を3バイトで表現可能として計算

3 ポーズ挿入規則

3.1 予備検討

従来規則に基づいてポーズ挿入された合成音声の聴取実験によって収集した悪い評価の文を、文の言語構造に着目して分析することによってポーズ挿入規則の精緻化を試みる。そのため、聴取実験においては、ベースとなるポーズ挿入規則で作った合成音声の各文の各文節間に挿入されたポーズの良し悪しを被験者に評価してもらい評価結果を蓄積する。

予備検討の結果、ポーズの挿入された個所に関して、ポーズの要不要を判定することはできるが、ポーズの挿入されていない個所に関して、ポーズに要不要を判定することは困難であることがわかった。必要なポーズが欠落している場合に、文全体が棒読みであるという印象を受けるが、どこかのポーズが不足しているかを判断することは困難である。そこで、以下の聴取実験において、合成音声の各ポーズ挿入個所に関して、ポーズの要不要を判定し、あわ

せて、文全体としてポーズが不足しているかを判断した。結果的に、文全体としてポーズが不足と評価された文はなかった。

3.2 実験方法

対象とする文：京大コーパスの1,000文を用いる。
そのうち、500文を用いて、係り受け距離のみによるポーズ制御の評価を行い、別の500文を用いて、制御規則を精緻化したものの評価を行う

被験者： 20代女性2名、20代男性1名

合成音声： 市販の音声合成システムを用いる。
京大コーパスの係り受け情報を用い、本ポーズ制御規則に基づき、音声合成システムから出力される韻律制御コードを変更して音声合成する。ヘッドホン受聴とし、文を見てもよいとした
次の2項目に関して官能評価する。

係り受け距離のみによるポーズ制御の合成音声：
合成音声のポーズ個所に関し、ポーズが必要（合成音声のままでよい）かポーズが不要（合成音声に含まれるポーズがない方がいい）かを判断（評価対象のポーズの総数は1477個所）

制御規則を精緻化した合成音声： 合成音声のポーズ個所に関し、ポーズが必要かポーズが不要かを判断（評価対象のポーズの総数1584個所）

3.3 実験結果 — 係り受け距離のみによるポーズ制御 —

ベースとなる制御方式として、係り受け距離のみから次のようにポーズ制御する。

- 係り受け距離1なら、ポーズなし
- 係り受け距離2なら、短いポーズ
- 係り受け距離3以上なら、長いポーズ

3名の被験者の判定結果を図6に示す。合成音声のポーズ個所に関し、ポーズが必要（合成音声のままでよい）か不要（合成音声に含まれているポーズがないほうがよい）かを判断してもらっているのので、ポーズ必要は制御がよいことを意味し、ポーズ不要は制御が悪いことを意味する。よって、満足の割合を $P/(P+Q+N)$ 、(P:ポーズ必要, Q:ポーズ不要,

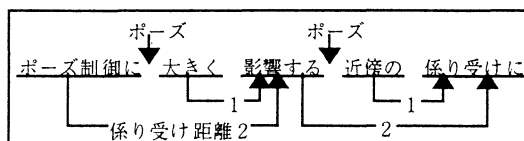


図6 係り受け距離によるポーズ制御例

N: どちらとも言えない) あるいは、 $(P+N)/(P+Q+N)$ 、で表すことができる。結果を図7左側に示す。3名全員の満足度は、前者の指標で84.1%、後者の指標で87.1%、相乗平均85.6%である。

また、被験者間の評価の重なる割合を調べると次の通りである。

全員Pとした数/いずれかがPとした数 = 68.1%

全員Qとした数/いずれかがQとした数 = 1.3%

3.4 実験結果 — 制御規則の精緻化 —

先の係り受け距離によるポーズ制御に、従来研究で言われていることを次の規則にまとめ、先の係り受け距離による規則に追加する。

- (1) 連続した複数文節の文節末にポーズがあり、それらのいずれの文節末にも句読点がない場合、初めの文節末にのみ長ポーズを入れ、その他の文節末に短ポーズを入れる
- (2) 連続した複数文節の文節末にポーズがあり、それらのいずれかの文節末に句読点がある場合、初めの句読点にのみ長ポーズを入れ、その他の文節末に短ポーズを入れる
- (3) 句読点と区切り記号「・」には、少なくとも短ポーズを入れる

実行順序は、(1)、(2)を並行して行い、その結果に対して(3)を適用する。結果を図7右側に示す。3名全員の満足度は、前者の指標で90.2%、後者の指標で92.0%、相乗平均91.1%である。また、被験者間の評価の重なる割合を調べると次の通りである。

全員Pとした数/いずれかがPとした数 = 81.9%

全員Qとした数/いずれかがQとした数 = 9.4%

4 まとめ

局所係り受けの精度は90%強の高い精度で、かつ、実装しやすいポーズ制御のための係り受け解析手法を開発した。本法は次の特徴を持つ。

- 保守や移植に有利な機械学習方式である決定リストを採用している
- ポーズ挿入位置決定の目的に十分な局所係り受け解析を採用している
- 処理アルゴリズムがシンプルなため、いろいろなシステムへ組み込む際の移植が低コスト
- 学会発表されている最高性能の手法であるME法に比べて、若干精度が低い、ME法にないメモリ容量と計算速度の可変性がある

次に、音声合成におけるポーズ挿入位置制御のための規則を作成し、聴取実験によって性能を確認した。この規則の中心となる主要因は係り受け距離で

あり、係り受け距離のみに基づく制御で、約86%のポーズ挿入位置が挿入適当という結果であった。さらに、句読点や、ポーズの連続などの要因を取り入れて規則の精緻化を行った。その結果、約91%のポーズ挿入位置が挿入適当という結果が得られた。

謝辞

プログラム、実験等でご協力をいただいた当研究所、菅原朋子殿、白木伸征殿に感謝致します。

参考文献

- [1] 内元清貴, 関根聡, 井佐原均: 最大エントロピー法に基づくモデルを用いた日本語係り受け解析, 情報処理学会論文誌, Vol.40, No.9 (1999)
- [2] D. Yarowsky: Decision lists for lexical ambiguity resolution: Application to Accent Restoration in Spanish and French, Proceedings of the 32th Annual Meeting of the Association for Computational Linguistics
- [3] 杉藤 美代子監修: アクセント・イントネーション・リズムとポーズ, 三省堂 (1997)
- [4] 海木延佳, 匂坂芳典: 局所的な句構造によるポーズ挿入規則の検討, 電子情報通新学会論文誌 D-II (1995)
- [5] 宇津呂武仁, 西岡山滋之, 藤尾正和, 松本 裕治: コーパスからの日本語従属係り受け選好情報の抽出およびその評価, 自然言語処理, Vol.6, No.7 (1999)

