

音声対話における頑健な主語補完手法

山本 和英 隅田 英一郎

ATR 音声翻訳通信研究所

E-mail: yamamoto@itl.atr.co.jp

6C-5

1. はじめに

日本語を対象にした音声対話処理においては、主語などの要素がしばしば省略されるため、これらを補完するための様々な研究が行われてきた。しかし実際に対話を処理する際に入力となるのは音声であり、部分的に誤りを含んだ音声認識結果が処理対象になるため、補完処理においても頑健性が要求される。このように頑健性を考慮した格要素補完、特に主語補完処理は音声対話処理実現のための重要な処理であるにもかかわらず、これまでほとんど研究が行われていない。

以上の動機に基づき、我々は以前に[Yam98]で提案した補完手法を基本にして、より頑健性を持ったモデルを構築し、実験によって頑健性を確認した。本稿では、提案手法の概略と実験結果を述べる。

2. 頑健性を持ったモデルの条件

ある問題、例えば省略された主語を求める問題に対して、解答の信頼度を向上させるための一つの方策は、複数の解答候補を用意してその中から選択することである。すなわち、解答に至るまでの手段を複数用意することで解答候補を得る、もしくは解答手段に利用する情報源つまり解の手がかりを別個にし、それぞれの情報源を別個に利用することで複数の解答候補を求め、その中から適当な解を選択することによって、頑健性を持たせることができる。

ここで、複数の解答候補が得られたときこのうちどれを最終的な解答とするかは重要な問題である。一つの方策として多数決モデルなどが考えられるが、この導入のためには各解答候補の信頼性が同等となるように解答候補を用意する必要がある。このため我々はこの方策を取らず、各解答候補の信頼性を計算し、最終的な解答を決定する際に信頼性が最大の候補を選択する。

3. 複数決定木 (MDT) モデル

前節の考察に基づき、要素の挿入、削除など、入力の部分的な不正確性に対する頑健性を持ったモデルを提案する。このモデルは、我々が[Yam98]で提案した格要素省略補完モデルを拡張したものであり、複数決定木モデルもしくはMDT (Multiple Decision Tree) モデルと呼ぶ。モデルの概要を図1に示す。

複数決定木モデルは単一の決定木モデル (Single Decision Tree, SDT) を基本にして、同一の学習事例集合に対して属性集合のみを変化させ、同一の決定木学習

表 1: 用意した属性集合

集合 (要素数)	内容語 200	機能語 166	話者情報 1	計 367
集合 A	内容語	+機能語	+話者情報	367
集合 C	内容語		+話者情報	201
集合 F		+機能語	+話者情報	167

方法によって複数の決定木を作成する。決定木の学習では、正解主語情報が付与された(誤りのない)事例を入力にして、事前に用意した各属性の有無によって質問を行ない、エントロピー基準によって事例集合の分類を行なっていく。

属性集合としては表1に示す3種類を用意した。各属性集合は、内容語群、機能語群、話者情報の組み合わせで構成される。話者情報は音声認識によって誤ることはないので、全属性集合に含めた。話者情報以外に関しては、入力形態素中のある探索範囲における当該形態素の有無によって属性の照合を行なった。

3.1 補完候補の選好基準

図1において、属性集合 S_1 による補完結果候補 D_1 よりも、属性集合 S_2 による補完結果候補 D_2 のほうが解の信頼性が高いと考えるのは自然である。なぜなら、両決定木は同一の学習事例集合から得られる木であるにもかかわらず、解候補 D_1 を得るための属性群と同一の特徴を持った学習事例数が D_2 のそれよりも少ないためである。このことは、解候補 D_1 を得るために使用した属性群の一部が比較的稀であり、すなわちそれらの一部が誤りである可能性が比較的高いことを示唆している。

以上の考察により、我々は決定木学習時の終端節点の事例数に着目した。すなわちMDTモデルではSDT学習時の各終端節点の事例数を信頼性とみなした。これにより、各SDTが出力した解答候補のうち、終端節点事例数が最大のSDTが出力した解答をMDTでの解答とする。例えば図1では属性集合 S_2 における解答の学習時終端節点事例数が最も多いので、 D_2 を解答とする。

4. 音声認識結果に対する性能試験

本稿で提案したモデルの有効性を確認するため、実際の音声認識結果を入力とした実誤りに対する補完精度を測定した。また比較のため、音声認識誤りのない正解入力に対する補完精度も測定した。訓練事例数は1401事例、実験事例数は訓練に含まれない303事例である。対象ドメインはホテルの予約もしくは解約時の二者会話であり、ATR旅行会話コーパスを使用した。

使用したMDTは、前述した3種類の属性集合(A,

Robust Subject Ellipsis Resolution Model in Spoken Dialogue.

Kazuhide YAMAMOTO and Eiichiro SUMITA

ATR Interpreting Telecommunications Research Labs.

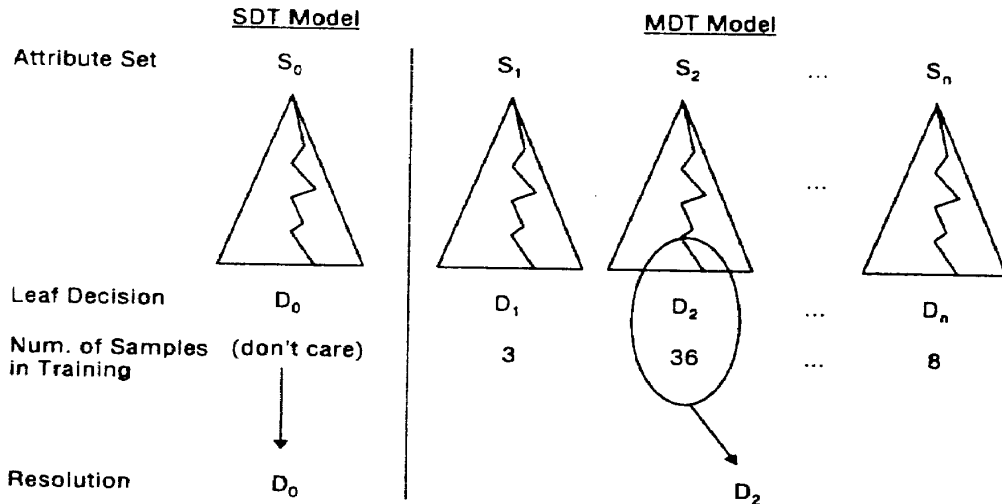


図 1: MDT モデルの概要

表 2: 音声認識誤りに対する主語補完精度

パラメータ	P1	P2	P3	(正解入力)
単語認識率 (%)	78.48	78.89	72.09	(100)
SDT/A	77.2	76.5	76.2	81.8
SDT/C	73.9	74.9	73.2	80.8
SDT/F	75.5	74.2	72.2	79.5
MDT	78.2	78.8	77.5	83.5

C, F) による SDT を使用した。すなわち、SDT/A、SDT/C、SDT/F が出した複数解答のうち、学習時に同一終端節点になった事例数が最も多かったものを MDT における補完結果とした。また比較のため、各 SDT 単独の補完結果も測定した。

音声認識装置は日英音声翻訳システム ATR-MATRIX における音声認識用音響・言語モデル [Nai98] を使用した。実験では、認識装置の音響尤度と言語尤度の相対的加重を変化させることによって 3 種類の異なる誤り傾向をもつ音声認識結果を用いて行なった。各パラメータの音声認識精度と主語補完精度を表 2 に示す。なお、表における補完精度は F 値 (%) [Yam98] によって示す。

表 2 により、実際の音声認識の結果誤りを含んだ入力に対して本稿で提案する MDT モデルが従来手法 (SDT/A) よりも全パラメータにおいて高精度であることがわかる。MDT は 3 種類の SDT が持つ情報のみを使用しているにもかかわらず、どの SDT よりも高い補完精度を持っている。このことは、3 種類の SDT が異なる結果を補完候補とした場合において、3.1 節で提案した候補選好基準が正しく機能していることがわかる。

4.1 正解入力に対する精度

本研究の目的は誤りを含んだ入力に対する頑健な処理であるが、仮に誤りがなくても MDT は有効に機能する可能性がある。なぜなら、相異なる情報源を手がかりに得られた解答候補に対して本稿の提案する選好基準によって解答を決める仕組みは、学習時に類似事例が最大であった属性群の解答を支持することを意味するためである。この仮説を検証したのが、表 2 の右欄「正解入力」である。これによると、正解入力に対しても MDT はどの SDT よりも明らかに高い補完性能を示すことがわかる。

5. まとめ

音声対話に頻出する主語省略の補完処理に関して、頑健性を持った複数決定木 (MDT) モデルによる補完手法を提案し、実験により手法の有効性を確認した。また提案手法が、誤りのない正解入力に対してもより高い補完精度であることを示した。今後は、人工誤りに対して実験することで、誤りの種類と頑健性の関係を考察する。

参考文献

- [Nai98] 内藤正樹, 政瀧浩和, SINGER, H., 塚田元, 匂坂芳典: 日英音声翻訳システム ATR-MATRIX における音声認識用音響・言語モデル, 春期講演論文集, 2-Q-20, 日本音響学会 (1998).
- [Yam98] YAMAMOTO, K. and SUMITA, E.: Feasibility Study for Ellipsis Resolution in Dialogues by Machine-Learning Technique, In *Proc. of COLING-ACL'98* (1998).