

対話の構造と単語の概念を利用した発話の理解

4C-1

荒木雅弘, 齊藤 隆, 佐藤研治, 西田豊明, 堂下修司

京都大学

1. はじめに

人間とコンピュータが音声による会話を行ないながら、何らかの問題を解決するシステム(音声理解システム)を構築するためにはどのような問題点があるのだろうか。我々は、次の3つを重要な問題と考える。

- 音声認識技術の高度化
- ノイズを含んだ入力に対する文解析手法の開発
- 対話処理を行なうことによる断片的な発話の理解

現在開発中の対話システムでは3つのサブシステム(音声認識部・文解析部・対話処理部)に分けて、それぞれの問題に対処する方法を探求している。本論文では主に文解析部と対話処理部の処理方式およびその統合法を述べる。

2. スケジュール管理に関する対話システムの概要

我々の研究室ではこれまで、音声認識や自然言語理解の研究を行ってきたが、それらの要素技術が将来の高度な知的通信サービスシステムの中でどのような役割を果たすかを評価するために、人間-機械系の対話システムを作成している[1]。タスクとしては個人およびグループ内のスケジュール管理を扱っている。

図1にこのスケジュール管理に関する対話システムの構成を示す。また、以下で現在開発中の3つのサブシステムの基本的な方針を述べる。

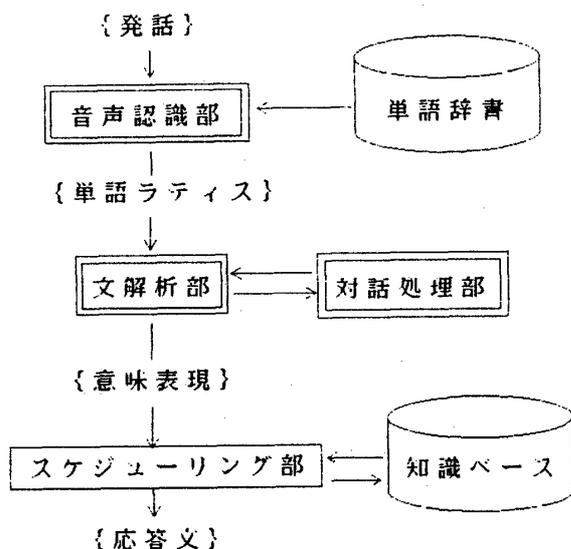


図1 スケジュール管理に関する対話システムの構成

2.1 音声認識部

不特定話者の自然な連続発生を高精度で認識することを目指す。認識の基本的な方針はまず、入力音声中から定常部を取り出し、そこを母音部、その間を子音部としてセグメンテーションを行なう。母音部・子音部それぞれに対して、2群対判別の手法を用いた音素片シンボル識別器を走査して音素片系列を得る。この音素片系列を入力パターンとしてHMM(Hidden Markov Model)による認識を行う[2]。なお、音素単位のHMMを結合することによって単語単位のHMMを作成し、それによって単語認識を行なう。音声認識部の出力は単語ラティスとする。

2.2 文解析部

音声入力による文を解析する際には、通常自然言語処理を行なうアルゴリズムをそのまま適用しても、うまくゆかないことが多い。それは、通常自然言語処理で生じる曖昧性に音声認識部で生じる曖昧性が加わり、処理すべき候補数の組合せ的爆発が起こってしまう。そこで、できるだけトップダウン的に意味候補を絞り込む手法として、単語ラティス中のスコアの高い重要単語数個から意味候補を絞り込んでゆく方法を提案する。手法の詳細は4章で述べる。

2.3 対話処理部

対話処理を行なう理由は主に2つある。1つは文解析部の意味候補絞り込みにおいて有効な予測情報を生成するためであり、もう1つは対話文に特徴的に現れる問題(省略表現・代用表現などの解釈、指示語の指示物の同定など)を解決するためである。ここではスケジューリングというタスクに特徴的に現れる対話構造のモデルを構築し、それを使って対話を処理する手法について述べる(3章)。

3. 対話構造のモデル化

人間は、談話にある定まった再現的なパターンを利用して、対話活動を行っていると思われる。ここで談話とは、話題が導入されてからその話題が放棄されるまでに行われる、結束性をもった発話の時間的系列のことを指す。以下において対話の構造を、談話状況と社会的行為を用いてモデル化することについて考察する。

目的のある対話では、依頼より始まる談話が多く現れると思われるので、ここでは、依頼より始まる談話に限って考察を進める。

談話内に位置づけられる発話の使用には規則性があり、ある発話がなされた後に行うことができる発話には制限がある。人間はこの制限に従って、発話を生成しかつ理解している。そこで、発話を社会的行為の観点から分類することを考える。依頼より始まる談話の構造モデルを図2に示す。

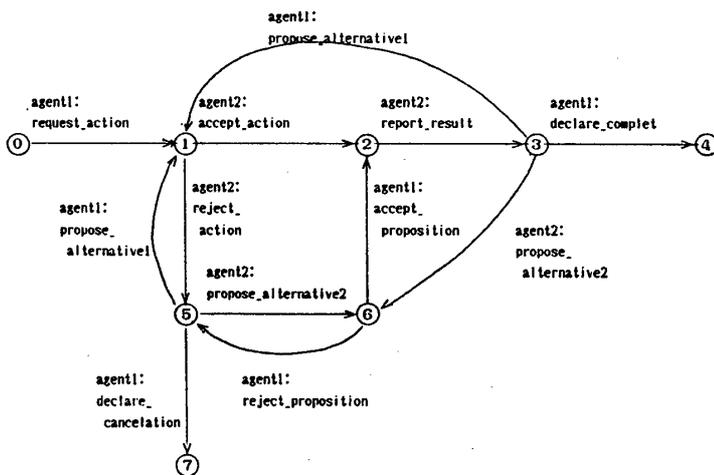


図2 依頼より始まる談話の構造モデル

ある社会的行為が聞き手に正しく識別されるためには、その行為が談話中の適切な場面で行われる必要がある。そこで社会的行為を行うのに適切な談話の場面を記述するために、8つの談話状況を設定する。それぞれの談話状況では、遷移先として可能な孤に対応する社会的行為を実行することができる。遷移先の談話状況が複数個あるのは、ある社会的行為が沈黙として表現された場合、ある談話状況が飛ばされて次の談話状況に移ったように見えるからである。

例えば、対話例1、対話例2の共にscene0において s p 1 が発話されたとする。対話例1では、s p 2 は scene1において reject_actionが行われているのに対して、対話例2では、s p 2 はscene2における report_resultが行われている。これは、対話例2では s p 2 が沈黙によって、scene1における accept_actionを実行してしまったからだと思われる。

●対話例1

s p 1 : 「明日の会議は何時からだっけ？」
(request_action)

s p 2 : 「知らない。」 (reject_action)

●対話例2

s p 1 : 「明日の会議は何時からだっけ？」
(request_action)

s p 2 : 「10時からです。」 (report_result)

ここで導入した談話状況から現在の発話に続き得る発話の内容(意味)を予測し、次に述べる文解析部に予測情報として渡す。

4. 単語の概念集合からの意味候補の選択

ここでは音声入力によって得られた単語ラティスに対して複数キーワードスポッティングを行ない、その結果得られた単語集合から意味候補を選ぶという方法について説明する。

意味候補を作成する時に用いる単語の概念は図3に示すようなネットワーク構造で表現している。この構造の中で対話処理部からの予測情報により次に発話され得る意味として適当なものに対して、複数キーワードスポッティングを行なう。つまり、可能な意味の全てについて、それぞれの意味が持つ概念に属する単語を単語ラティスの中で探し、重なりがないように単語が得られればその組合せを持つ意味を正しいとして選ぶことになる。

この方法では結果として複数の意味が出てくるが、選んだ単語の単語ラティス中でのスコアを掛け合わせたものを、その意味のスコアとして出すようにして、スコアが最大のものを解析結果とする。

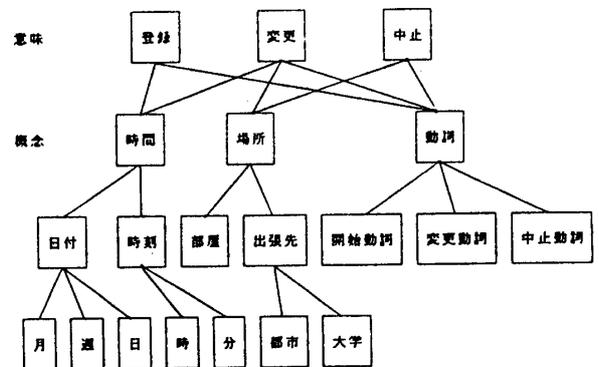


図3 単語の概念ネットワーク

5. おわりに

今回発表した文解析部と対話処理部は別々に、SUN-4上のSicstus-Prologで現在開発中である。今後は各サブシステムが完成後、順次結合して手法の有効性を評価してゆきたい。

参考文献

[1] 荒木他：概念情報を利用した会話文の構造解析、第4回人工知能学会全国大会(1990)。
[2] 河原他：判別分析とHMMの統合による不特定話者子音認識、信学論 D-11 Vol. J73-D-11 No.9 pp1363-1372(1990)。