

プラン認識による発話認識結果の選択

8B-9

月岡陽一 秋葉友良 伊藤克亘 田中穂積
(東京工業大学)

1 はじめに

対話における相手の発話を音声認識する際、その認識結果は統語解析や意味解析では判別できず、語用論的な解析をしなければ選択できないような音韻的に類似した候補が生じる [1]。例として以下のような対話例を考える。

visitor :	音声研究会に出席する予定ですが、行き方がわかりません。
system :	音声研究会は、どこで開かれるのですか。
visitor :	機械振興会館です。
system :	機械振興会館なら行ったことがあります。

visitor :	(1) とうとうふうに行けばいいですか。
	(2) とうとうふうに行けばいいですか。
	(3) とうとうふうに行けばいいです。

これは、visitor が、system から音声研究会に参加するための情報を聞き出そうとしている対話の一部である。この対話においては、音声認識によって visitor の最後の発話として (1) から (3) のように三つの認識結果が得られる。これらの候補は意味解析まで施されており、単独の文としては誤りがない。しかし、文脈の流れから判断すると不自然な候補を含んでいることがわかる。ここでは、(2) の「とうとうふうに行けばいいですか。」が尤もらしい認識結果として選択されるべきである。

このような選択を行なうには、語用論的な解析が必要である。本研究では、このように一つの発話を音声認識する際に生じる複数の発話入力候補から、プラン認識の手法を利用して文脈的に正しい候補を選択することを目標とし、システムを構築した。

2 システムの概要

本システムは、音声認識部 [2] とプラン認識部から成る (図 1 参照)。visitor の発話は音声認識部で音韻列として認識され、さらに構文・意味解析されて意味構造として出力される。我々はこの意味構造を発話入力候補と呼ぶ。この発話入力候補は、独立した文としては正しいが、語用論解析を施さなければ文脈的な妥当性を判定できない候補として複数生成されている。プラン認識部はこのように複数の発話入力候補を入力として、語用論的レベルからもっとも妥当な候補を選出する。

A method for selecting the pragmatically well-formed candidate of speech recognizer using plan-recognition technique
Yoichi TSUKIOKA, Tomoyosi AKIBA,
Katunobu ITOU, Hozumi TANAKA
Tokyo Institute of Technology

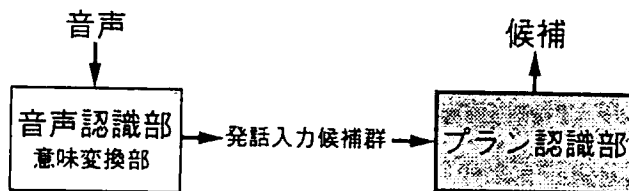


図 1: システムの概略

3 プラン認識

プラン認識とは、観測された相手の発話行為をもとに、相手を持っているプランを推定することである。プラン認識に関する従来の研究としては、Allen らの研究 [3] があげられる。そこでは、プラン認識を利用して、相手の発話では直接表現されていない相手のプランや目標を推定することで、より協調的な応答を実現することを提案している。

Allen らのプラン認識の手法は、発話入力から複数のゴールに向かって推論する前向きプランニングと、逆にゴールから推論する後向きプランニングの二種類を併用してプランニングを行なう。

プランニングとは、あらかじめ用意されたオペレータ (基本行為) を組み立て、あるゴール状態を満たすようなオペレータの系列を見つけることである。オペレータには、それが表す行為が実行されるための preassumption (前提条件) と、その行為によって引き起こされる effect (効果) が定義されている。プランニングは、あるオペレータに対してその effect を、preassumption として持っているオペレータを新たに接続する手続きを繰り返すことによって実現される (図 2 参照)。

4 プランナの拡張

Allen らの研究では、観察される相手の発話行為には曖昧性がないことを仮定している。また、相手の発話に対する応答を生成することを目的としており、そのためには相手のプラン、ゴールを完全に認識する必要がある。

それに対し本研究では、音声認識からの認識を行なうため、相手の発話行為は複数の候補として得られる。プラン認識自体がもともと計算量が膨大な上、複数の誤認識された発話入力候補を開始行為としてそれぞれにプランニングを施さなければならないため、何らかの方法で計算量の削減を試みなければならない。また音声認識部からは既に音韻的な情報をもとにスコアリングされた候補

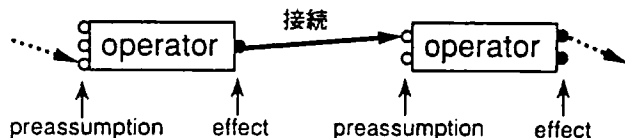


図 2: オペレータの接続

が生成される。よって音声認識部からのスコアと、プラン認識の尤もらしさから得られるスコアを統合する必要がある。さらに、相手のプランを組み立てることは候補選択の手段にすぎないため、プランを完全に認識する必要はない。

本研究で用いるプランナは以上のような相違点を考慮して、以下に述べるヒューリスティクスを導入して、Allen のプランナを拡張した。(図 3 参照)

4.1 プランの長さに関するヒューリスティクス

複数の発話入力候補すべてを偏ることなくプランニングするために、一つのオペレータを接続する毎に任意の割合でスコアを低くする操作を行なう。

4.2 ゴールまでの距離に関するヒューリスティクス

プランニングは、接続するオペレータによって生じる effect を、システムが管理している信念に反映させて、その信念がゴール状態をすべて満たした時に終了する。そこで、システムの信念がゴール状態にどれだけ接近しているかをスコアに反映させて、できるだけゴール状態に近いプラン系列を先に処理するヒューリスティクスを導入する。

4.3 オペレータの接続に関するヒューリスティクス

プラン系列の中には、以下にあげるようなオペレータの接続例がしばしば現れる。

- (1) request(visitor, system, inform(system, visitor, fact) inform(system, visitor, fact)
- (2) request(visitor, system, action) inform(system, visitor, visitor want action)

(1) は、system が visitor にある事柄を教えるように visitor が system に提案した (request) 後、system が visitor にその事柄を教える (inform) ことを意味していて、妥当性の高いオペレータの接続である。一方 (2) は、visitor が system にある行為を提案した後、オウム返しに system が visitor に visitor はその行為を欲していることを伝えることを意味していて、部分的なオペレータの接続関係は成立しているが、ある目標を達成するためのプランとしては無意味なものである。このように隣接するオペレータの接続については局所的な妥当性が評価できるため、それをメタ知識として保持してスコアの操作を行なう。

4.4 後向きプランニングの省略に関するヒューリスティクス

オペレータの preassumption は複数存在するため、前向きプランニングにおいて新しくオペレータを接続するには、接続に関与した以外の preassumption が成立していなければその接続は認められない。そこで接続するオペレータの preassumption が満たされない時には、それを満たすプランを新たに生成するために後向きプランニングが実行される。(図 3 参照)

しかしこの後向きプランニングをすべて実行するのは、非常にコストがかかる。さらに、ある種の preassumption は、相手の信念の誤り、system の知識の不足、要求など不確実な要素、などを含むために、しばしばプランニン

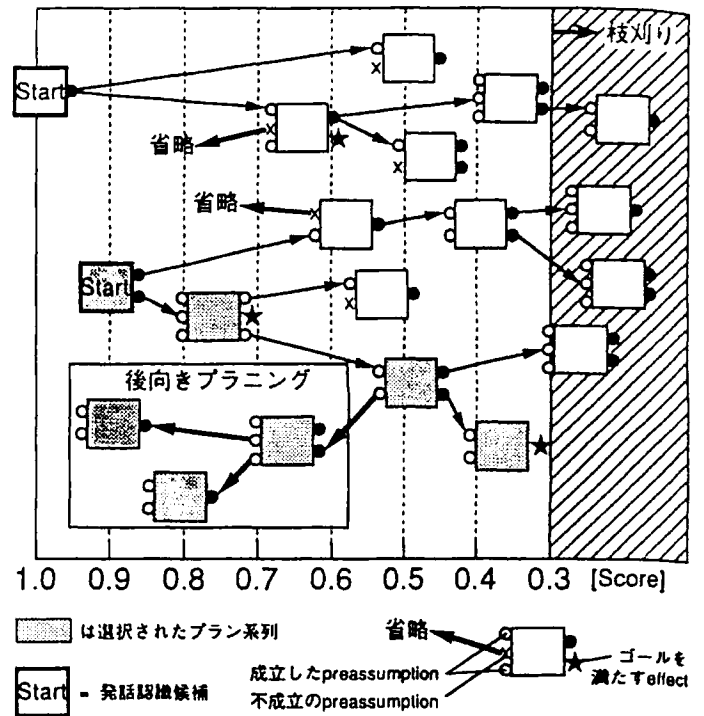


図 3: プランニングの例

グ不可能な場合がある。そこで重要性を基準にオペレータの preassumption を分類し、後向きプランニングを実行する必要度が低い preassumption については、後向きプランニングを省略してスコアを下げるだけで先に進むようにした。その結果、処理量が減少する上、system と visitor 間の、信念のズレや誤解などが原因で生じるプランニングの失敗に対する扱いを、統一的に処理することができる。

5 まとめ

本研究では、プラン認識を利用した発話認識結果を選択する手法と、そのために有効なプランナを提案した。現段階では、システムの持つ信念知識や相手のゴールに対する情報は、既知なものとして初期状態で与えられている。今後の課題としては、相手の発話認識結果をもとに、システムが持つ信念知識や相手のゴールを更新することによって、これらの情報をインクリメンタルに構築しなければならない。

参考文献

- [1] 永田昌明. 統計的な対話モデルの試みとその音声認識への応用. 人工知能学会研究会資料, No. SIG-SLUD, pp. 94-103, Jul 1992.
- [2] 伊藤克巨, 速水悟, 田中穂積. 音素文脈依存モデルと高速な探索手法を用いた連続音声認識. 電子情報通信学会論文誌, Vol. J57-D-II, No. 6, pp. 1023-1030, Jul 1992.
- [3] J. F. Allen and C. R. Perrault. Analyzing intention in utterances. AI, Vol. 15, No. 3, pp. 143-178, 1980.