

連続音声認識からの概念表現の作成

6D-2

野口淳 奥村明俊 畑崎香一郎 坂井信輔 渡辺隆夫
(日本電気(株) C&C 情報研究所)

1. はじめに

音声入力を自動通訳システム、音声QAシステム等に使用する場合は入力音声の一言一句を明らかにするのが主な目的ではなく、むしろ入力音声の意味を取り出すことすなわち入力音声の理解が必要となる。

従来、これらのシステムを実現するためには、入力された音声を生声認識によって文字列に変換した後、入力音声の意味を抽出していた。しかしこの方法を用いると、解析のための多くの計算量や辞書情報が必要とされ、また解析誤りも生じやすい。

入力された音声の意味を解析せずに抽出する方式としては文献[1]がある。この方式では、タスクを限定し、音声認識用のネットワーク上の各単語に意味情報を付与し認識結果として、解析を行なうことなしにその意味構造を出力する。

一方、文の意味を表す枠組として機械翻訳等で用いられている中間言語[2]がある。この中間言語はタスクに依存せずに記述できるので、入力音声の意味を表すものとして用いることによっていろいろなタスクに容易に対応できる。

本稿では、有限状態オートマトン等のネットワークを用いた連続音声認識において、ネットワーク内の各単語間の依存関係や素性情報等をあらかじめ記憶しておくことにより認識結果から中間表現を作成する方式を提案する。この方式を用いることにより、認識結果から構文解析、意味解析を改めて行なう必要がなくなり、かつ種々のタスクに対応することができるようになる。また、一言一句解析するのではないので文の意味を変えない範囲内であれば言い回しの自由度を増やすことが可能である。

2. 全体のながれ

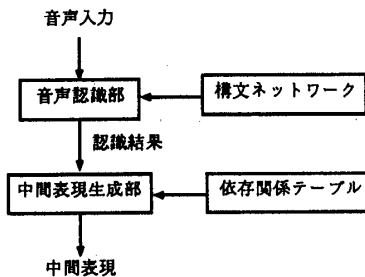


図 1: 構成図

図 1 に本方式の構成図を示す。音声認識部ではネットワーク制御の連続音声認識[3]を行なう。音声認識の結果

* Interlingua output from Continuous Speech Recognition, by Jun Noguchi, Akitoshi Okumura, Kaichiro Hatazaki, Shinsuke Sakai, and Takao Watanabe (C&C Information Technology Research Laboratories, NEC Corp.)

果、ネットワーク上のもっとも尤度の高い経路が認識結果として中間表現作成部に出力される。依存関係テーブルにはあらかじめ認識結果に格関係や素性情報を与えるための情報を用意しておく。中間表現生成部は、音声認識結果と依存関係テーブルの情報とから中間表現を作成し出力する。

連続音声認識用ネットワーク

音声認識に用いるネットワークの一例を図 4 に示す。ネットワーク上の各アークが単語を表している。一部のアークではサブネットワークが呼び出される。Start から End までの経路のうち、入力音声をもっとも良く照合する経路が認識結果として出力される。

中間言語

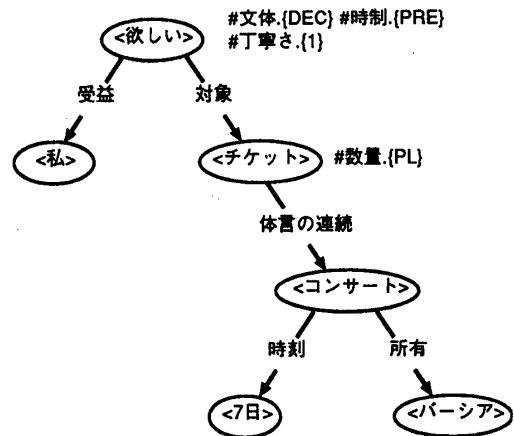


図 2: “7日のパーシアのコンサートのチケットが欲しい”

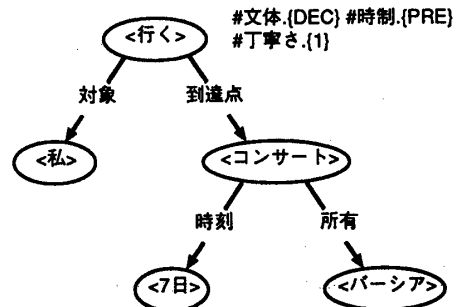


図 3: “7日のパーシアのコンサートに行きたい”

中間言語は自然言語の一文単位の意味を表層言語の種類の依存しない形式で表すものであり、言語独立な概念素

(CP)をノードとし概念素間の依存関係をアークとする有向ネットワークとして表現されている。

図4のネットワークから生成される中間表現の例を図2、3に示す。〈コンサート〉、〈欲しい〉等は、それぞれ単語“コンサート”、“欲しい”の概念素である。図2では〈欲しい〉の受益格が〈私〉で、対象格が〈チケット〉であることを表している。

中間表現の各ノードには、文の時制、丁寧さ、数量等を与える素性情報が付与されている。

中間表現の生成

依存関係テーブルには中間表現の生成に必要な以下の情報が記述されている。

1. 概念素

認識用ネットワーク中の各単語に対する意味を表す概念素。

必要な場合には、入力音声に含まれない概念素を補う。図4のネットワークでは“私”という単語は省略されているが、図2、3に示したように中間表現を生成するためには“〈私〉”という概念素を補わなければならない。音声入力される際に省略されるがそれに相当する概念素が中間表現内で必要不可欠であればその概念素を依存関係テーブル中にあらかじめ定義しておき、中間表現を作成する際に補う。

2. 概念素間の依存関係

認識用ネットワーク中のアーク間の中間表現における依存関係。

3. 素性情報

中間表現における依存関係と概念素に付与すべき素性情報。

素性情報は他の概念素に影響されることがある。例えば“1枚の切符”と発話されたとき、中間表現内で概念素〈切符〉に付与される数量の素性情報は#数量.{SIN}であるが、“2枚の切符”と発話されたときは、概念素〈切符〉には素性情報#数量.{PL}が付与される。このような現象に対処するため、あるネットワークのアーク上の

単語が認識された場合、他のアーク上の単語に対応する概念素の素性情報に必要な情報を付与するようにあらかじめ依存関係テーブルに記述しておく。

中間表現生成部は、依存関係テーブル中の概念素のうち音声認識結果に含まれるものを取り出し、テーブル内に記述されている概念素間の依存関係に従って有向ネットワークを構成する。同時にテーブル内に記憶されている素性情報を概念素および依存関係に付与する。

3. おわりに

本稿では連続音声認識用の有限状態オートマトン等のネットワーク中の単語間の意味的な依存関係や素性情報等をあらかじめ記憶しておくことにより、連続音声を認識した結果として中間表現を出力する方式を提案した。

この方式により、入力音声の認識結果に対し構文解析、意味解析を改めて行なう必要がなくなり、かつ種々のタスクに容易に対応することができるようになる。また、文の意味を変えない範囲内での自由な言い回しを許すことも可能となる。

これまでに、実際にチケット予約タスク文の音声認識に本方式を適応した結果、有効性を確認した。

謝辞

本方式のプログラム開発を支援していただいた日本電気技術情報システム開発(株)の落合尚良氏に感謝致します。

参考文献

- [1] 小林, 白井, 「ネットワークモデルによる会話音声理解における焦点の表現法」, 日本音響学会音声研究会資料SP85-15、1985
- [2] 市山, 村木, 「機械翻訳システムPIVOTの中間言語」, 情報処理学会第38回全国大会、1989
- [3] 服部, 塚田, 吉田, 渡辺, 「半音節HMMによる連続音声認識」, 電子情報通信学会技術報告SP89-15、1989

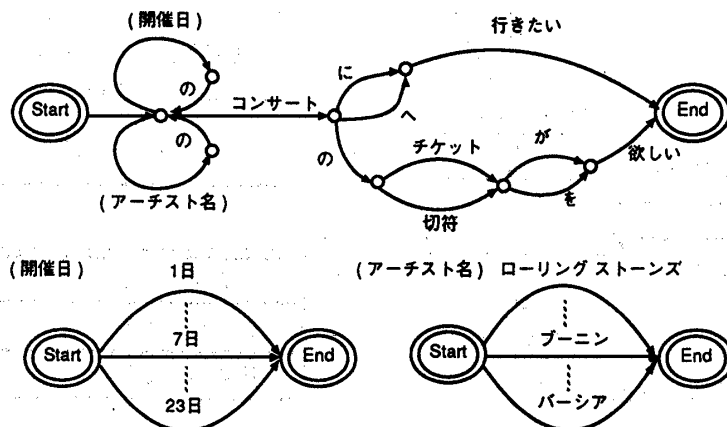


図4: 連続音声認識用ネットワーク