論文 / 著書情報
Article / Book Information

論題(和文)	WFST駆動の音声認識を用いた乗換案内システムにおける駅名の統計情報の利用
Title(English)	
著者(和文)	Novak Josef, Whittaker Edward W. D., 古井 貞熙
Authors(English)	JOSEF R NOVAK, Edward Whittaker, SADAOKI FURUI
出典(和文)	日本音響学会2009年秋季講演論文集, , No. 3-1-17, pp. 129-130
Citation(English)	, , No. 3-1-17, pp. 129-130
発行日 / Pub. date	2009, 9

WFST駆動の音声認識を用いた乗換案内システムにおける駅名の統計情報の利用*

◎Novak Josef, Whittaker Edward W. D., 古井 貞熙 (東工大)

1 はじめに

本論文では、iPhone向け、WFST駆動の音声認識を用いた乗換案内システムにおける、駅名の統計情報を利用した実験の結果を報告する。特に、東京都内の1089個の固有の駅名で構成されたネットワークを用いて、駅名に対するユニグラム、また共起頻度の統計情報を利用したネットワークの性能を、重み付けなしのネットワークと比較する。また、日本全国の9283個の固有の駅名で構成されたネットワークを用いて、ユニグラムと重み付け無しのネットワークの比較を行う。

本論文では、まず2.にて、関連している文献をまとめる。次に、3.にて、WFSTを利用した音声認識の概要を述べる。4.にて実験設定・装置を述べる。5.にて、実験結果およびその解析結果を述べる。6.にて、まとめを述べる。

2 関連する文献

現在我々が開発している、音声認識を活用した検索システムは、国内の乗換案内に関連する情報を対象にしている。本分野は注目を集めており、特に[1]では、テレフォニー駆動のドイツの市内乗換案内に関するクエリーを受け付けることが出来るシステムが発表されている。我々も、[2]で、別のテレフォニー駆動の国内の乗換案内システムを発表している。

3 WFSTの音声認識

本研究の音声認識実験では、重み付け有限状態トランスデューサー (WFST) を使っている。WFSTのアプローチでは、様々な知識資源を比較的簡単に、かつ効率的に一つの探索ネットワークに合成できる。そして探索ネットワークの最適化によって、高速、かつ高精度の音声認識ができる。以上の理由により、近年WFSTのアプローチが注目を集めている.

この論文で発表する実験では、三つのWFSTを合成することによって探索ネットワークを構築する。具体的には、$C \circ L \circ G$ ネットワークを構築する。ここで、C は文脈依存音素から文脈非依存音素へのWFST、L は文脈非依存音素から単語へのWFST、そして G は単語から文法へのWFSTである。

WFSTの詳細は、[3]を参照のこと。

4 実験の設定・装置

本論文で利用したシステムは、クライアント・サーバー方式であり、クライアントは、iPhone 3G向けのアプリである。録音はクライアントで行い、音声認識の処理はサーバー側で行う。

実験では、音響モデルは日本語話し言葉コーパス (CSJ)[4]を学習データとして、CMUのSphinxTrainによって8kHzのモデルを学習している。認識文法は、記述文法で約１００個の録音された乗換案内の要求文を用いて、人手で記述されたものである。この中で最も使われた、“出発・到着時間”、“終電”、“始発”、“明日”などを含む乗換案内についての質問の言い回しを、文法に含めた。以前に発表したシステムと違い、本システムでは、キーワードなどの順序の制約は一切しない。例えば次のような言い方もできる。

「六本木から上野駅まで、07:30出発」、

「午後10:00までに到着、恵比寿から広尾駅まで」。

この実験に使用した駅名の統計情報は、株式会社ぐるなび[5]によって提供されたものである。モデルを構築するため、またそれに対する様々なディスカウント手法の効果を評価するため、CMU Language Modeling Toolkit [6]を利用した。

実験に使用したテストセットは、iPhoneで録音した476個の音声クエリーである。

5 実験とその結果

ここでは、駅名の統計情報を利用した様々なネットワークを比較した結果について述べる。

5.1 ディスカウント手法について

様々なディスカウント手法が認識精度にどのような影響をするか検討するため、Linearディスカウント、Absoluteディスカウント、Good-Turingディスカウント、Witten-Bellディスカウントという四つディスカウント手法に基づいて、共起頻度情報を用いたモデルを構築した。Table 1では、この四つのモデルにおける駅名の認識精度を比較する。

この結果によると、ディスカウント手法によって、駅名の認識精度が大きく変わることはないが、以下

*Application of Station Co-occurrence Statistics to a WFST-based ASR Train Timetables System. by NOVAK, Josef, WHITTAKER, Edward W. D. and FURUI, Sadaoki (Tokyo Institute of Technology)

	G.T.	Lin.	Abs.	W.B.
% accy	87.3	87.5	87.5	87.7

Table 1　駅と駅の共起頻度情報の様々なディスカウント手法における認識精度。ここでは、G.T.: Good-Turing、Lin.: Linear、Abs.: Absolute、W.B.: Witten-Bell。

の実験では、比較的性能が高かった Witten-Bell ディスカウント手法を利用する。

5.2　認識精度の実験結果

本実験では、二つの一般探索ネットワークを用いた。このネットワークは、含まれている駅名の数以外は、全く同じ構造でできている。一つは、*Tokyo-to*（東京都）という、1083 の東京都内にある駅でできたもの、もう一つは、*Zenkoku* という日本全国に及ぶ 9283 個の駅名を含んだ探索ネットワークである。*Tokyo-to* の探索ネットワークにおける実験の結果を、Fig. 1 で表す。ここでは、ユニグラムや共起頻度情報を用いた探索ネットワークが、重み付け無しの場合と比較されている。

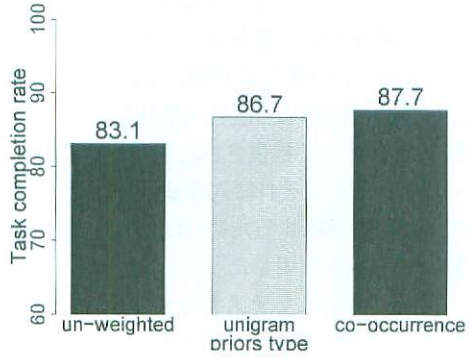

Fig. 1　Station recognition accuracy for the Tokyo-to network.

この結果によると、共起頻度情報を利用した探索ネットワークは、ユニグラムを用いた探索ネットワークより少し精度は高いが、共起頻度情報を用いた探索ネットワークは、ユニグラム情報を利用した探索ネットワークの約 35 倍の資源を必要とする。

従って、*Zenkoku* では、共起頻度情報を含んだ探索ネットワークを構築しないことにした。Table 2 に結果を示す。ユニグラム統計情報を用いた探索ネットワークは、重み付け無しのネットワークより大きく精度が上がっている。共起頻度情報を含めても、*Tokyo-to* と同様に、精度がユニグラムより大きく上がらない可能性が高く、その代わりに、n-best の候補を使用した方がユーザにとっては価値があると考えられる。

	un-weighted	unigram
% accy	65.2	84.4

Table 2　*Zenkoku* に対する駅名における認識精度

6　まとめ

本論文では、我々が開発している WFST 駆動の音声認識を用いた iPhone 向け乗換案内システムの最新結果を報告した。特に、東京都に限ったネットワークを用いて、駅名に対するユニグラム、また共起頻度の統計情報を利用したネットワークの性能を、重み無しのネットワークと比較した。また、日本全国のネットワークを用いてユニグラムと重み無しのネットワークの比較を行った結果も報告した。*Tokyo-to* の結果によって、共起頻度情報を利用した探索ネットワークは、ユニグラム情報のみを使用したものより精度が大きく上がらないため、*Zenkoku* ではユニグラム情報のみ使用した。将来は、共起頻度情報の他にも、n-best 結果を利用した実験や、さらにより柔軟な n-gram を用いた言語モデルを、今まで使用してきた記述文法と比較したいと考えている。

謝辞　データを提供して頂いた株式会社ぐるなび、また Inferret Japan 株式会社に感謝する。

参考文献

[1] W. Eckert, et. al., "A Spoken Dialogue System for German Intercity Train Timetable Inquiries," Proc. European Conf. on Speech Technology, pp. 1871-1874, 1993.

[2] E.W.D. Whittaker, et. al., "A Prototype Spoken Natural Language Interface for Information Access on Mobile Phones," In Proc. ASJ, 2007.

[3] M. Mohri, et. al., "Weighted finite-state transducers in speech recognition," CSL, vol. 16, no. 1, pp. 69-88, 2002.

[4] K. Maekawa, "Corpus of Spontaneous Japanese Its Design and Evaluation," Proc ISCA, pp. 7-12, 2003.

[5] http://www.gnavi.jp

[6] http://www.speech.cs.cmu.edu/SLM/toolkit.html