G_014

中規模語彙を対象とした音声認識システム用のFSA言語モデルの自動獲得

Automatic Cosntructon of FSA Language Model for Middle-sized Vocabulary Speech Recognition System

森元　逞†　　高橋　伸弥†　　吉原　龍市‡
Tsuyoshi Morimoto Shin'ya Takahashi Ryuichi Yoshihara

1．まえがき

1,000 語程度の中規模語彙を対象とした音声認識システム用の言語モデルの 1 つとして，FSA 言語モデルがある．この言語モデルを直接書き下すのは非常な手間が必要となる．このため，正規言語クラスの文法定義ツールを用意し，記述された文法から FSA に自動変換できるようにしたシステムもある（例えば，文献[1]の HParse）．しかしそれでも充分かつ矛盾の無い文法を定義するのは大変である．一方，大量の学習データが得られる環境であれば，バイグラム（bi-gram)やトライグラム（tri-gram)などの統計言語モデルが有効である．しかし得られる学習データ量があまり多くない場合は，統計情報の信頼性が低くなり，結果として性能の良くない言語モデルになってしまう恐れがある．なおこれまでも，与えられた学習データから FSA を自動的に構築する方法がいくつか提案されている[5][6]が，これらも統計量を用いた方法であるため，学習データ量が不十分であれば，同様な問題を有することになる．

本論文では，日本語会話文を対象とし，学習データの DP マッチングを行なうことにより FSA 言語モデルを自動的に構築する手法を提案する．また旅行会話例文を対象とした FSA の構築と，その FSA を用いた音声認識実験の結果について報告する．

2．手法の概要

(1) DP マッチングに先立ち，例文をそれら相互の距離によりグループ分け（クラスタリング）する．これは，あまり共通点のない例文同士を無駄に DP マッチングしてしまうことを避けるためである．文間の距離としては，以下のように共通する単語数を用いて計算する．またクラスタリング手法としては最大距離アルゴリズムを用いる．

$$\text{距離}= 1-\frac{(\text{例文 A の単語}\cap\text{例文 B の単語})}{(\text{例文 A の単語}\cup\text{例文 B の単語})}$$

(2) 各クラスタ内の例文の DP マッチングを行なう．まず適当な 2 文を取り出し，両者それぞれを，単語をノード，単語間の接続をリンクとする FSA に変換し，それぞれを x 軸，y 軸に配置してノード同士の DP マッチングを行なう．これにより，ノード間の「一致」，「置換」，「挿入」，「削除」の関係を求め，これらの関係に基づいて y 軸の FSA に x 軸のノードをマージする．さらに，次の 1 文を取り出して x 軸に配置し，また上記の FSA を y 軸に配置して DP マッチングを行ない，結果を y 軸の FSA にマージする．

†福岡大学工学部電子情報工学科
‡福岡大学大学院工学研究科

以上の処理を例文数だけ繰り返す．なお，FSA との DP マッチングでは，FSA のリンクに基づいてマッチングすべきパスを決定する．

3. DP マッチングと FSA の作成

3.1 DP マッチング

本手法で用いているDPマッチングは通常のDPマッチング・アルゴリズムとほぼ同じであるが，どのパスでマッチングを行なうかは，FSAのリンクに従うものとする．また同表記・異品詞語を区別するため，単語情報として，表記と品詞名の両方を用いる．なお品詞名としては「茶筌」[3]で定義されている品詞名を用いている．

$g(x,y)$:ポイント(x,y)におけるグローバル距 離
$d(x,y)$:ポイント(x,y)におけるローカル距離
Δy:FSAリンクにおける yの 1 つ前のポイントの y座標

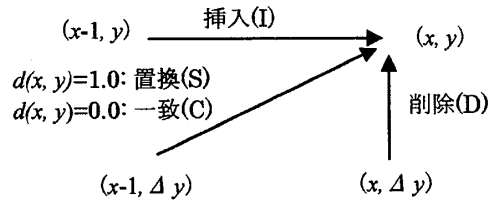
$$g(x,y)=\min\begin{pmatrix} g(x-1,y) \\ g(x-1,\Delta y) \\ g(x,\Delta y)\end{pmatrix}+d(x,y)$$

where

$$d(x,y)=\begin{cases}1.0 & (w_x\neq w_y)\\ 0.0 & (w_x=w_y)\end{cases}$$

3.2 FSA の作成

DP マッチングの結果，ポイント(x, y)におけるローカル距離と，そのポイントにおける 1 つ前のポイントが求まるが，これによりノード同士の関係を以下のように決定する．



次に，これらの結果に基づき，前述したように y 軸上の FSA にノードのマージを行なう．

4．音声認識実験と評価

4.1 コーパス

市販されている旅行英語会話例文集 4 冊を対象とし，そこに記載されている日本語例文を収集した．ただし，あまりにもくだけた口語表現は除外した．また文数の少なさを補うため，収集した例文を参考として作成した文を追加し，

合計 1,000 文のコーパスを用意した．本コーパスにおける異なり語彙数は 1,254 である．

4.2 クローズド・データに対する実験

上記 1,000 文のコーパスを学習データとし，クラスタ数を変化させて FSA を作成した．また学習データ中からランダムに 60 文を選び，音声認識を行なった．音声認識に用いたシステムの諸元を表 1 にしめす．

表 1　音声認識実験システムの諸元

音声認識システム	HVite[1]
HMM	4 混合性別非依存のトライフォン[2]
音声データ	60 発話（男性話者 3 名が異なる文を発話）

構築された FSA の平均分岐数（リンク数／ノード数）ならびに音声認識結果を表 2 にしめす．また，比較のために同一コーパスに対するバイグラム言語モデルのパープレキシティおよび音声認識結果を合わせてしめしている．

表 2　平均分岐数ならびに音声認識実験結果（クローズドデータ）

クラスタ数	平均分岐数	単語認識率 (%)	文認識率 (%)
30	1.45	98.7	90.0
50	1.42	98.7	90.0
70	1.40	98.9	91.7
バイグラム	5.88（パープレキシティ）	93.0	70.0

本手法により構築された FSA は平均分岐数がかなり小さいこと，またその結果，音声認識性能が極めて高く，特に文認識率がバイグラムよりも 20%ほど高くなっていることことが分かる．なお，クラスタ数の増加にともなって平均分岐数は多少減少し，その結果音声認識率もわずかに向上している．

4.3 オープン・データに対する実験

オープンデータに対する性能を評価するため，音声データ 60 発話に対応する文を学習データから除外した 940 文から FSA を作成した．なお，元々の学習データ数が絶対的に少ないために，60 文を単純に削減すると，語彙自身が FSA から削除されてしまう恐れがある．このため，まず学習データ内に現れる名詞については意味素性に基づく単語クラスに置き換えて DP マッチングを行って一旦 FSA を作成し，次にその FSA における単語クラスのノードを所属する単語群に展開して最終的な FSA を作成することとした．意味素性を求めるために，「分類語彙表」[4]を使用し，その分類コードの少数点以下の 2 桁を用いた．

空港：	1.	2	6	4	0
	品詞	人間活動の主体	社会・機関	公共機関	公共機関の場所

結果を表 3 にしめす．クローズドデータに比し，音声認識率，特に文認識率が大きく低下している．これは，FSA から名詞以外の単語（特に元々出現頻度が 1〜2 程度であったもの）が削除されたこと，実験対象発話に対応するパスが削除されたこと，などが原因であると思われる．それでもバイグラムに比べると，まだ性能はかなり良いと言える．

表 3　平均分岐数ならびに音声認識実験結果（オープンデータ）

クラスタ数	平均分岐数	単語認識率 (%)	文認識率 (%)
30	1.70	78.8	26.7
50	1.67	79.9	21.7
70	1.68	79.4	28.3
バイグラム*)	6.36	58.4	3.3

*) バイグラム言語モデルでは，学習データから 60 文を削除せず，かわりにバックオフ（ウイットン・ベル）を有効とした．

4.4 受理可能な文数

作成された FSA は，オープンデータに対する性能は必ずしも満足できるものではないが，一方ではノードの共用と単語クラスの導入により，元の学習データより多くの文を受理できることは確かである．どの程度の文を受理できるかを評価するため，作成された FSA を用いてランダムに 500 文を生成し，そのなかで日本語として正しい文がどの程度あるかを目視でチェックした．ただし，正しいかどうかの判断は，構文・意味的な判断だけでなく，その文が旅行会話で用いられそうな文かどうかも考慮した．なお，対象とした FSA は 4.3 で述べた 940 文，クラスタ 70 で生成したものである．結果を表 4 に示す．

表 4　生成された文の正否（重複を除く）

	文数
学習データと同一	75
正しい	97
多少おかしい	18
誤り	123

この結果より，学習データの文数を 75 と仮定すると，学習文以外に 97 の正しい文を生成しているから，元の学習データに対し，

$$\frac{(75+97)}{75} = 2.29$$

倍の文を受理できるものと評価できる．

5. まとめ

学習データの DP マッチングを行なうことにより FSA 言語モデルを自動的に構築する手法を提案した．また旅行会話例文を対象として FSA を構築し，その FSA を用いた音声認識実験結果をしめした．オープンデータに対する音声認識性能は必ずしも満足できるものではないが，バイグラムよりはかなり良い性能を達成できている．今後はさらにオープンデータに対する性能の向上方法について検討を進める予定である．

【参考文献】
[1] S.Young, et al.: The HTK Book Ver.3.0
[2] 河原，他：連続音声コンソーシアムの活動報告および最終版ソフトウエアの概要，情処研究会報告，SLP-49-57, 2003
[3] 松本，他：日本語形態素解析システム「茶筌」V.2.0 使用説明書，NAIST Tech. Rep, IS-TR99012, 1999
[4] 国立国語研究所：「分類語彙表」形式による分類語彙表（増補版），1996
[5] 山本，他，：HMM を言語モデルに用いた連続音声認識の検討，情処第 45 回全大, Vol. 3, 1992
[6]J. Hu, et al.: Language Modeling with Stochastic Automata, Proc. of ICSLP'96, 1996