小型音声合成エンジンPureTalkの日本語解析*

4Q−2

廣田 誠　相澤道雄　久保山英生　八木沢津義
キヤノン(株) プラットフォーム技術開発センター†

1　はじめに

カーナビや各種モバイル機器をはじめとする近年の情報機器の多様化，あるいは，社会のバリアフリー化の進展を背景として，音声合成技術への期待が急速に高まっている．情報機器の多様化 (Non-PC 化) の流れを考えた場合，搭載する音声合成エンジンの小型化が重要な要件となる．我々が開発している音声合成エンジン PureTalk[1] は，すでに小型で高音質な音響処理モジュールを実現している．一方，今後の音声合成の用途として，電子メール，Web 情報など，任意のテキストを読み上げるアプリケーションやサービスが広く求められると予想され，小型で高精度な言語処理モジュールが求められる．本研究では，PureTalk の言語処理モジュールとしての利用を目的とした，小型で高精度な日本語解析エンジンを開発した．本稿では，その中心となる形態素解析の内容について説明する．

2　形態素解析の基本モデル

形態素解析の基本モデルとして，「クラス接続コスト最小法」を採用した．まず，単語を，直前の単語との接続属性 (前接属性) に基づいて定義した前接クラスと，直後の単語との接続属性 (後接属性) に基づいて定義した後接クラスに分類する．そして，後接クラスと前接クラスの間の接続のしにくさを接続コストという数値で表現することにより単語列の尤もらしさを評価する．このモデルは，以下の式で計算される総コストを最小にする単語列を求めるものである．

$$Cost = \sum_{i=1}^{N+1} C(t_k^{i-1}, t_z^i) + \sum_{i=1}^{N} L(W_i|t_z^i)$$

ここで，t_z^n，t_k^n はそれぞれ，n 番目の単語の前接クラス，後接クラスである．$C(t_k^{i-1}, t_z^i)$ はクラス間接続コストであり，タグ付きコーパスから計算できる連接確率の log を適当な範囲の整数に変換することにより求める．一方，$L(W_i|t_z^i)$ は単語 W_i の単語コストであり，W_i が属する前接クラス内での出現確率の log を適当な範囲の整数に変換することにより求める．このモデルは，統計的言語モデルの一つであるクラス bigram([2]) と等価である．最適解は，整数和によるコスト計算と DP により求まるため，比較的小さなプログラムで実現できる．

3　クラスの定義

クラス bigram モデルの性能を左右する大きな要因の一つは，クラスの分類方法である．クラスの分類方法については，統計的なクラスタリング手法を用いる方法が提案されている ([2][3])．これに対し，我々は，従来から人手の作業によって開発してきた単語のさまざまな属性情報を前接属性，後接属性として利用することで，以下のようなクラス分類を行なった．

- 前接クラスは，品詞と前接属性によって分類する．
- 後接クラスは，活用型と後接属性によって分類する．

名詞の前接属性，後接属性の例を表 1 に示す．

表 1: 名詞の接続属性例

前接属性例

固有名詞後接性	0：固有名詞に後接しにくい 1:人名に後接しやすい (ex) 弁護士 2:地名に後接しやすい (ex) 駅
後接複合性	0:後接複合しにくい (ex) 昨日 1:後接複合しやすい
数詞後接性	0:数詞に後接しにくい 1:数詞に後接しやすい (ex) 部屋，大会

後接属性例

副詞性	0:副詞的でない 1:副詞的である (ex) 昨夜
前接複合性	0:前接複合しにくい (ex) 次 (つぎ) 1:前接複合しやすい

こうしたクラス分類の方法は，小型で高精度な言語処理モジュールの実現にとって，以下のような長所がある．

*A Japanese language analysis module for small-sized Text-to-Speech engine "PureTalk"

†Makoto Hirota, Michio Aizawa, Hideo Kuboyama, Tsuyoshi Yagisawa(Canon Inc. Platform Technology Development Center)

- 人間の先見的言語知識の利用という従来的手法と統計的手法のメリットを効果的に融合でき，高精度化につながる．
- クラスから，品詞，活用型などの基本的な文法情報が一意に定まる．従って，解析辞書の各単語には，クラスのID情報と，単語コストを付与すればよい(音声合成の場合，これ以外に読みやアクセントの情報が必要)ので解析辞書サイズの小型化につながる．

4 解析辞書の圧縮

日本語解析エンジンを小型化するのに最も重要な要素は，解析辞書の小型化である．本研究における解析辞書小型化のポイントは以下のとおりである．

- 前述のクラス定義による付与情報の種類削減．
- 単漢字辞書の各漢字の読みは，最も頻度の高い1つだけに限定する．
- 上記単漢字辞書を使って表記から読みを復元できる単語については，読み情報を付与しない．
- 各単語に付与する情報はハフマン符号化により圧縮する．
- 収録語数を単語の出現頻度をベースに調整する．

5 接続コストの誤り訂正学習

前述のように，クラス間接続コストは，タグ付きコーパスから計算できる連接確率から自動的に求める．しかし，こうして求められた接続コストが解析精度の面から最適な値になるとは限らない．そこで，解析誤りをなるべく減らす方向へ接続コストをチューニングするために，タグ付きコーパスを用いた誤り訂正学習を行った．学習アルゴリズムは以下のとおりである．

1. タグ付きコーパス中の原文すべてを形態素解析する．
2. 解析結果とタグ付きコーパスの正解とを比較し，不一致の単語列を双方から抜き出す．
3. 解析結果中の不一致単語列について各単語間のクラス接続コストを $(1+\alpha)$ 倍する $(0<\alpha\ll 1)$．
4. 正解中の不一致単語列について各単語間のクラス接続コストを $1/(1+\alpha)$ 倍する．
5. 全文に対する精度を計算する．過去の精度と比較して値の変化がある程度小さくなったら終了．でなければ1へ戻る．

6 性能評価

以上に説明した言語処理モジュールとすでに開発している音響処理モジュール全体でのサイズと解析精度の評価を行なった．サイズは，ROMサイズ(=解析辞書，波形辞書，プログラムの合計サイズ)，RAMサイズ(ヒープ使用量)を測定した．解析精度は，新聞記事を対象とし，単語分割，品詞，読みトータルの適合率，再現率を評価した．なお，解析辞書の収録語数や波形辞書の設定を変え，より小型に絞ったケースと，フルスペックのケースについて評価した．ROMサイズと解析精度は，表2のようになった．RAMサイズは入力文長に依存するが，標準的な日本語文であれば500Kバイトで十分おさまることがわかった．

表2: 性能評価結果

	小型ケース	フルスペック
解析辞書語数	4万語	28万語
波形辞書の設定	11KHz, 圧縮	22KHz, 非圧縮
解析辞書サイズ	650(Kbyte)	3300(Kbyte)
波形辞書サイズ	250(Kbyte)	1500(Kbyte)
プログラムサイズ	300(Kbyte)	
ROM 合計	1200(Kbyte)	5100(Kbyte)
解析精度 適合率	97.43(%)	98.31(%)
解析精度 再現率	97.94(%)	98.39(%)

7 おわりに

小型音声合成エンジンPureTalkの言語処理モジュールについて説明した．品詞，活用型などの基本形態素情報，先見的言語知識を反映した各種属性情報を前接クラス，後接クラスという形に一元化し，クラス接続コスト最小法というシンプルなアルゴリズムを採用することで，高精度で小型な言語処理モジュールを実現した．性能評価の結果，ROM1Mバイト，RAM500Kバイト程度のリソースで，精度の高いテキスト音声合成が可能であることがわかった．

参考文献

[1] M.Yamada, et al., "PureTalk: A high quality Japanese text-to-speech system", Proc. ICSLP 2000,pp.403-406,2000.
[2] Brown P.F, Pietra V.J.D, deSouza P.V, Lai J.C, and Mercer R.L, "Class-Based n-gram Models of Natural Language", Computational Linguistics, 18(4),467-479, 1992.
[3] 山本博史，匂坂芳典，"接続の方向性を考慮した多重クラス複合N-gram言語モデル"，信学技法，NLC98-38,49-54,1998.