数詞句アクセントを考慮した CRF による日本語のアクセント型変形予測モデルの改良*

©小林俊平, 清水信哉, 峯松信明, 広瀬啓吉 (東大)

1 はじめに

任意の日本語テキストを入力としてその自然な読上げ音声を出力する日本語テキスト音声合成 (Text to Speech) を実現するためには, その文中のアクセント句境界位置, 及び, 各句内のアクセント核位置を適切に推定する必要がある. 日本語のアクセント処理を難しくしている一つの要因として, 孤立単語を発声した場合と複数単語を連続発声した場合とでそのアクセント核の位置が変化する「アクセント結合」が起こることが挙げられる. よってより自然な音声を機械に出力させるためには, 文発声時の「アクセント句境界」と, 文中に現れる自立語の孤立発声時のアクセント核やその他の種々の言語情報などから, 与えられた文に対して, 適切なアクセント結合を推定するモジュールの開発が必須である.

このようなアクセント処理に関する代表的な先行研究として, アクセント価や結合様式などの属性を定義することで規則によりアクセント結合を記述した句坂らの研究 [1] があり, 従来アクセント処理に関しては, この手法が広く用いられてきた. 一方で筆者らは, CRF(条件付き確率場) と比較的大規模なアクセントラベルが施されたコーパスを用いた, 統計的なアクセント結合処理手法を提案し [2], それに上記のアクセント結合規則を考慮した素性を加えることで [3], 単純な規則処理の手法よりも高い精度を実現している. しかしどの先行研究においても, 全てのアクセント句に対して適切な処理をできているとは言いがたい. エラー解析を試みると, 特に数詞を含む句や外来語を含む句など, 特殊なアクセント変化を起こす句で誤りが多いという問題があり, 素性を改良するなどして, これらの句における予測精度向上が求められている.

そこで本稿では, 上述した CRF を用いた既存手法において特に誤推定率の高かった数詞を含むアクセント句に対し, 数詞特有の特性を考慮した素性を追加することで, 精度を更に上げることを試みる. 具体的には, 数詞のアクセント規則を網羅的に記述した宮崎 [4] の先行研究を参考に, その規則を CRF の学習素性として使える形に変換して追加し, 推定精度の向上を試みた.

本稿の構成は以下のようになっている. まず 2 節では先行研究におけるアクセント推定精度とエラー解析の結果を述べる. 次に 3 節で数詞句に関する先行研究と提案手法で用いる新たな素性について説明する. さらに 4 節と 5 節で実験結果とその考察を述

Table. 1 Features used for training CRF in the previous method[2, 3]

<ul style="list-style-type: none"> ・観測素性 (当該形態素および前後 2 形態素のラベルを利用) -[基本形, 基本形読み, 書字形, 品詞, 活用型の組み合わせ] -[品詞] -[品詞 (大分類のみ)] -[活用型 (大分類のみ)] -[活用形 (大分類のみ)] -[モーラ数] -[単独発声アクセント型] -[当該形態素と隣接形態素の組み合わせ素性] -[単独型種類ラベル] -[アクセント結合様式] -[結合アクセント価]
--

べ, 最後に 6 節で本稿をまとめる.

2 アクセント結合予測に関する先行研究 [1, 2, 3]

2.1 アクセントデータベース

筆者らはまず, 新聞記事読上げコーパス (JNAS)[5] で使用されている文 (毎日新聞記事 12,516 文) に対し, 以下のようなラベル付けを行ったコーパスを作成した [6].

- 文発声時の「アクセント句境界」と「アクセント核位置」
- 文中に現れる自立語の単独発声時の「アクセント核位置」

これらのラベル付けは, 一人のラベラによってなされている. ラベラを一人にした理由としては, 複数人によるラベリングでは広く知られているように, アクセント感覚の個人差 (東京方言話者であったとしてもアクセント感覚にはずれがある) が混入することを避けるためである.

2.2 先行研究におけるアクセント予測の実装とその評価

2.2.1 実験条件

先行研究で構築されたアクセント結合予測手法及びその精度について概説する. ただし実験には, 2.1 で述べたアクセントコーパス 12,516 文 (46,086 アクセント句, 131,467 形態素) から, 学習用として 10,684 文 (38,900 アクセント句, 114,783 形態素) を用い,

*Improvement of CRF-based accent sandhi prediction by considering accentual rules of numeral phrases. by KOBAYASHI, Shumpei, SHIMIZU, Shinya, MINEMATSU, Nobuaki, HIROSE, Keikichi (The University of Tokyo)

Table. 2 Results of accent type estimation in the previous study

	総数	規則ベース		CRF ベース		
		正答数	正答率	正答数	正答率	
形態素	16,682	14,615	87.6%	15,936	95.6%	
アクセント句	全ての句	7,184	6,171	85.9%	6,796	94.6%
	単純な句	2,287	2,126	93.0%	2,174	95.1%
	名詞連続	1,000	866	86.6%	958	95.8%

Table. 3 Error analysis of the results obtained in the previous study

	全ての句	単純な句	名詞連続	数詞	外来語	付属語連続
該当アクセント句数	7,184	2,287	1,000	751	316	5,190
誤答数	338	113	42	73	292	287
正答率	94.6%	95.1%	95.8%	90.3%	92.4%	94.5%

残りの1,832文(7,184アクセント句, 16,682形態素)を評価用に用いた。また, CRFによる処理を行うツールキットとしてはCRF++[7]を使用し, 結果の集計は形態素単位とアクセント句単位で行なっている。また本実験では, 正解に複数アクセント核がある場合¹, 二つ目以降の核は副次アクセント核であると考え, 一つ目のアクセント核の位置のみに着目して正誤判定を行なった。一方, 正解が無アクセントの場合は, 推定結果も無アクセントと判定された場合を正解とした。

さらに先行研究では, 頻繁に現れる特徴的な句として, 下記の2種のアクセント句に着目している。

● 単純なアクセント句

{ 名詞, 動詞, 形容詞, 形状詞 } + { 助詞, 助動詞 } の2語で構成されたもの

● 名詞連続を含むアクセント句

2語以上連続して名詞が出現するアクセント句

また, CRFの学習素性に利用した主な素性はTable.1の通りである。なお, 素性の詳細な情報や, アクセント結合規則に関する情報はそれぞれ[2, 3]と[1]を参照していただきたい。

2.2.2 結果

Table.2に, 1節で述べた句坂らの規則に基づく手法と, CRFを用いた統計的手法の推定結果を示す。

Table.2より, CRFを用いた統計的手法のほうが, 規則を用いる手法よりもすべての項目において高い精度を示している。単純なアクセント句や名詞連続を含む句のような頻繁に現れるアクセント句に対しては, 95%を超える高い精度を実現することが出来ている。

2.3 エラー解析

次に2.2.2で示したCRFにおける手法の実験結果に対して, 先行研究[8]で行なっているエラー解析の結果がTable.3である。

この結果より, 特に正答率の低いのは数詞を含むアクセント句(90.3%)と, 外来語を含むアクセント句(92.4%)であることが分かる。次節以降では, この2つのアクセント句のうち, 最も精度の低かった数詞を含む句に焦点を当てることとする。

3 提案手法: 数詞句を考慮したアクセント処理

本節では, 新たな提案手法として数詞句のアクセントを考慮したCRFによるアクセント処理手法について説明する。

3.1 語頭・語末変化結合型

まずは数詞の読みに関する重要な要素である, 語頭・語末変化結合型[9]について説明する。日本語においては, 例えば「ホン(本)」という単語が「ボン」や「ポン」に変わるように, 語頭音や語末音が頻繁に変化する。オンラインで配布されている電子化辞書Unidic[9]では, このような変化の規則を記述するため, 語頭変化形という属性を設けている。同様に「イチ(一)」という単語が「イツ」に変わるような現象を扱うため, 語末変化形という属性を定義している。そして語頭変化形の決定に際しては前接要素が, 語末変化形の決定に際しては後続要素が何であるかが影響を与える。たとえば, 「ホン(本)」が濁音形を取るのには, 前接要素が「三」の場合であり, 「イチ(一)」が促音形を取るのには, 後続要素が「本」「階」「杯」などの場合である。このような隣接要素に対する影響力を記すために, 語頭変化結合型(iConType)・語末変化結合型(fConType)という属性を設けている。語頭変化結合型は, 前部要素が後部要素に与える規則であり, 語末変化結合型は, 後部要素が前部要素に与える規則である。

これらの語頭変化結合型と語末変化結合型が深く影響するのは数詞と助数詞がほとんどであり, これは数詞・助数詞のアクセントを決める上で非常に有効であると考えられる(現在無償で配布されているテキスト音声合成システムGalateaTalk[10]でも, 実際に

¹本研究ではアクセント句の中に複数の核を認めている

Table. 4 Classification of auxiliary numerals by their accent changes

T_k	例
α	個, 位, 時, 分(ふん), 時間, 歳, 羽, 通り, 斤, 層, アール, センチ, キロ, ドル, 度(ど: 温度, 角度), 階, 球, 巡, 乗, 週, 人前, 敗, 着(到着), 度目, 代目, 貫目, 幕目, 日目, 球目, 丁目, 晝, ケ月
β	間, 台, 軒, 票, 町, 艘, 代, 枚, 名, 面, 本, 枚, 丁
γ	升
δ	年(ねん), 段(階段), 番
ϵ	貫, 版, 銭, 回, 点, 巻
ζ	尺, 着(衣服), 角
η	円
θ	曲, 石(こく), 匹, 冊, 足, 拍, 脚, 局, 発
ι	合
κ	度(ど: 回数)
λ	人
μ	月(ががつ), 日(にち)
ν	寸

この2つの結合型を用いて、数詞句の読みとアクセント型を規則により決定している)。

よって語頭・語末変化結合型をうまく CRF の素性として利用できれば、数詞句のアクセント推定精度を上げるのに貢献できると考えられる。具体的には、促音のモーラにはアクセント核はこない、などという特徴をうまく表せると考えられる。

3.2 数詞句アクセントに関する先行研究

次に、数詞句のアクセントに関する先行研究として、宮崎 [4] の「数詞-助数詞アクセント変化表」について説明する。宮崎は、数詞の音韻変化や数詞-助数詞のアクセント変化に従って数詞を 19 個、助数詞を 13 個に分類し、そのそれぞれのグループにおけるアクセント変化の規則を示した。助数詞の分類は Table.4, 「数詞-助数詞アクセント変化表」は Table.5 である。なお、Table.4 と Table.5 の縦軸のギリシャ数字はそれぞれが対応しており、Table.5 の表中の 0~3 の数字はそれぞれ、

$$A(S_i, T_k) = \begin{cases} 0: \text{アクセント結合規則に従う} \\ 1: 0 \text{ 型となる} \\ 2: \text{助数詞の第一音節にアクセント核が移動する} \\ 3: \text{助数詞の最終音節にアクセント核が移動する} \end{cases}$$

を表している。

宮崎によると、これらの規則は数詞のみに当てはまる特殊な規則であり、先行研究である句坂らによるアクセント結合規則にはこの規則は適応されていない。

3.3 数詞句を考慮した素性の追加

3.1 と 3.2 で示した二つの数詞に関する重要な要素を、従来手法で用いていた CRF の素性に追加する形で利用し、数詞句における精度向上を試みる。

3.3.1 語頭・語末変化結合型に関する素性

2.1 で示したコーパスには語頭・語末変化結合型の情報は載っていないため、形態素解析機 Mecab[11] と形態論辞書 Unidic[9] を用いて再度形態素解析し、これらの情報を得た。CRF 素性としての用い方としては、語頭・語末変化結合型が前後の形態素との規則であることを踏まえ、

- 当該形態素の語頭変化結合型と後続形態素の語末変化結合型の組合せ素性
- 当該形態素の語末変化結合型と前出形態素の語頭変化結合型の組合せ素性

として用いた。

3.3.2 数詞-助数詞アクセント変化表に関する素性

数詞-助数詞アクセント変化表に関する素性としては、Table.5 の横軸である“○(ゼロ)”から“幾”までの 19 個の数詞の分類 (以下“数詞カテゴリ”と呼ぶ) と、縦軸である“ α ”から“ ν ”までの 13 個の分類 (以下“助数詞カテゴリ”と呼ぶ) を用いて、以下のような素性を作成した。

- 当該形態素の数詞カテゴリと、前後形態素 (2 形態素まで) の数詞カテゴリの組合せ素性
- 当該形態素の数詞カテゴリと、前後形態素 (2 形態素まで) の助数詞カテゴリの組合せ素性

前者は数詞が連続した場合に考慮したものであり、後者は数詞と助数詞が連続して現れる場合に考慮したものである。

4 評価実験

2.2.1 で述べた実験条件と同じ条件で、提案手法の評価実験を行った。比較する手法としては、1: 句坂らによる規則ベースの手法、2: 句坂らの手法に宮崎の数詞-助数詞アクセント変化表の規則を取り入れた手法、3: 従来の CRF による手法、そして 4: CRF 手法に 3 節で示した数詞句アクセントに関する素性を加えた提案手法、の 4 つである。実験結果を、Table.6 に示す。

5 考察

Table.6 より、提案手法である、数詞素性を新たに加えた CRF 手法が最も良い精度を示すことが分かる。特に数詞を含む句へのエラー削減率は、従来の規則ベース手法と比べて約 72.0%、従来の CRF と比べても 37.0% を達成している。句坂らのアクセント結合規則に宮崎の数詞規則を取り入れた手法も、従来の

Table. 5 Accentuation rules of numerals-auxiliary numerals

$T_k \backslash S_l$	○	一	二	三	四	五	六	七	八	九	十	百	千	万	億	兆	数	何	幾
α	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0
β	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0
γ	0	0	0	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0
δ	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0	0
ϵ	2	2	2	2	0	2	2	0	2	0	0 ⁱⁱ	1	1	1	1	0	0	0	
ζ	3	3	3	0	0	3	3	0	3	0	0 ⁱⁱ	1	1	1	1	0	0	0	
η	0	1	1	1	0	0	1	0	1	0	0 ⁱⁱ	1	1	1	1	0	0	0	
θ	0	3	0	0	0	0	3	0	3	0	0 ⁱⁱ	1	1	1	1	0	0	0	
ι	0	2	2	0	0	2	2	0	0	0	0	1	1	1	1	0	0	0	
κ	0	3	3	0	3	3	0	0	0	0	0	1	1	1	1	0	0	0	
λ	0	0	0	2	2	2	0	0	0	2	0	1	1	1	1	0	0	0	
μ	0	3	3	0	3	0	3	3	3	0	3	3	1	1	1	1	0	0	0
ν	0	2	2	0	0	2	2	0	2	0	2	2	1	1	1	1	0	0	0

Table. 6 Result of accent type estimation of the proposed method

	総数	1: 規則 (従来)		2: 規則 (+数詞規則)		3: CRF(従来)		4: CRF(+数詞素性)		
		正答数	正答率	正答数	正答率	正答数	正答率	正答数	正答率	
形態素	16,682	14,615	87.6%	14,686	88.0%	15,936	95.6%	16,003	95.9%	
アクセント句	全ての句	7,184	6,171	85.9%	6,185	86.1%	6,796	94.6%	6,833	95.1%
	単純な句	2,287	2,126	93.0%	2,123	92.8%	2,174	95.1%	2,177	95.2%
	名詞連続	1,000	866	86.6%	870	87.0%	958	95.8%	960	96.0%
	数詞句	751	587	78.1%	604	80.4%	678	90.3%	705	93.9%

規則ベースの手法と比べれば約 10.4%程度のエラー削減率が得られたが、提案手法には及ばなかった。これは、宮崎の手法が数詞と助数詞が連続した場合にしか適応されないため、その効果が限定的であったことが原因であったためと考えられる。またもう一つ注目すべき点は、提案手法は数詞句だけではなく、他のアクセント句に対しても精度改善をもたらしている点である(数詞句以外のアクセント句でも従来の CRF 手法に比べ約 3.2%のエラー削減率を達成)。これは数詞のような特殊なアクセント句を考慮した素性を入れたことで、他の素性に対する重み付けも変わり、他のアクセント句に対しても良い影響を与えたことによるものと考えられる。

6 まとめと今後の課題

本稿では、日本語テキスト音声合成のためのアクセント推定に関して従来の CRF を用いた手法で誤推定率の高かった数詞を含むアクセント句に対し、その特徴を考慮した素性を組み込む手法を提案した。その結果、従来手法よりも数詞句に関して約 37.0%~72.0%のエラー削減率を達成した。また全てのアクセント句に対する推定精度も従来の CRF 手法より 3.2%向上させることができ、本手法の有効性が示された。

今後の課題としては、誤推定を起こしていた数詞句に関して詳細なエラー解析を行うとともに、さらに素性の改善を行っていくこと、そして外来語を含む句のような誤推定の多かった他のアクセント句に対

ⁱⁱ S_l の一つ前の形態素 $S_{l-1} \neq$ 二~九の時は規則が変化する。詳しくは [4] 参照

してもそれを考慮した素性の導入をしていくことなどが挙げられる。

また、アクセント結合処理技術の他分野への応用も考えている。ひとつの具体的例として、日本語のアクセント教育において、本手法で検討したアクセント変形予測モデルの導入を考えており、現在準備を進めている状況である [12]。

参考文献

- [1] 句坂芳典ら。日本語単語連鎖のアクセント規則。信学論, Vol. 66, pp. 849-856, 1983.
- [2] R.Kuroiwa et al. CRF-based statistical learning of Japanese accent sandhi for developing Japanese text-to-speech synthesis systems. *Proc. ISCA Workshop on Speech Synthesis*, pp. 148-153, 2007.
- [3] 印南圭祐ら。CRF を用いたアクセント変形予測モデルの規則処理に基づく改良。第 15 回言語処理学会年次大会発表論文集, pp. 574-577, 2009.
- [4] 宮崎正弘。日本文音声変換のための数詞読み規則。情報処理学会論文誌, Vol. 25, No. 6, pp. 1035-1043, 1984.
- [5] K.Ito et al. JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research. *Journal of the Acoustical Society of Japan*, Vol. 20, No. 3, pp. 199-206, 1999.
- [6] 黒岩龍ら。単独ラベラによる大規模アクセントデータベースの構築およびそれを用いた統計的アクセント結合処理の検討。電子情報通信学会音声研究会, pp. 31-36, 2007.
- [7] <http://crfpp.sourceforge.net/>.
- [8] 印南圭祐ら。CRF に基づくアクセント変形予測モデルにおけるエラー解析。第 14 回言語処理学会年次大会発表論文集, pp. 969-972, 2008.
- [9] <http://www.tokuteicorpus.jp/dist/>.
- [10] <http://sourceforge.jp/projects/galateatalk/>.
- [11] <http://mecab.sourceforge.net/>.
- [12] S.Kobayashi et al. Automatic generation of accent dictionary of conjugational words for any japanese text. *Proc. Int. Conference on Japanese Language Education*, pp. 784-786, 2011.