日本語音声合成を目的としたアクセント処理のための規則と統計的学習*

◎黒岩 龍（東大・情報理工）, 峯松信明（東大・新領域）, 広瀬啓吉 (東大・情報理工)

1 はじめに

日本語テキスト音声合成では、入力されたテキストに対し、まず付属する形態素辞書によるテキスト解析（形態素解析等）によって形態素境界や各形態素の読み（モーラ列もしくは音素列）・基本アクセント型などの情報を得、それを基に音韻処理（無声化などを含む処理）と韻律処理（アクセント・イントネーション・継続長・パワーなどの処理）がそれぞれ行なわれ、最終的な音声出力となる。辞書から得られる形態素単独発声時のアクセント情報は、文としてのアクセントとは異なるため、後の韻律処理において的確な処理（アクセント結合）をする必要がある。

この現象を規則によって表現する研究がなされており [1]、日本語テキスト音声合成システムとしての実装も行われている [2]。しかし、人手で与える規則で処理を行なう場合、すべての事象を網羅することは困難であり、例外的処理への対応も難しい。また、[1] ではアクセント句には高々1 つのアクセント核しか存在しないことを前提とするが、副次アクセントの存在を指摘する研究もあり、必ずしも適切なモデルではない。統計的手法としては、N-gram による形態素解析とアクセント処理の同時推定に関する研究がなされている [3]。

本研究では、条件付確率場を用いた統計的なアクセント処理について検討した。

2 条件付確率場を用いたアクセント処理の学習・推定

2.1 条件付確率場

近年、注目されている統計的学習モデルに条件付確率場（CRF; Conditional Random Fields）[4] がある。観測データ $\boldsymbol{x}$ を与えられた場合の出力ラベル $\boldsymbol{y}$ を学習するにあたり、CRF は $(\boldsymbol{x}, \boldsymbol{y})$ 内での連続する変数の組（y_{t-1} と y_t , y_t と x_t など）の関係についての独立した特徴（素性）f を列挙し、そのそれ各素性 f の重要度を θ_f、$(\boldsymbol{x}, \boldsymbol{y})$ 内で素性 f が満たされている箇所の数を $\phi_f(\boldsymbol{x}, \boldsymbol{y})$ とおいた上で、入力 $\boldsymbol{x}$ に出力 $\boldsymbol{y}$ を割り当てることの確信度合いとして、$\sum_f \theta_f \phi_f(\boldsymbol{x}, \boldsymbol{y})$ を考え、これを、

$$\Pr(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp \sum_f \theta_f \phi_f(\boldsymbol{x}, \boldsymbol{y})}{\sum_{\boldsymbol{y} \in Y} (\exp \sum_f \theta_f \phi_f(\boldsymbol{x}, \boldsymbol{y}))}$$

として確率分布とする。正解データを与えることによる学習は、この確率をできるだけ大きくするような重要度 θ_f を探る作業となる。

本研究では、CRF による処理は CRF++[5] を使用して行なった。

2.2 学習・推定の内容

本研究では、アクセント句境界は事前に与えられているものとし、各々のアクセント句の中でのアクセント核位置について学習と推定をした。文にアクセント句境界とアクセント核のラベルが付与されたテキストコーパスとして、黒岩らによるアクセントデータベース [6] の完成分 4,166 文を使用した。形態素解析結果や自立語単独発声時のアクセントについても、同データベースのものを使用した。

	A	B	C	D	E	F	G	H	I	J	K
1	まだ	マダ	未だ	マダ	副詞	*	*	マ'ダ／	2	1	1
2											
3	正式	セイシキ	正式	セイシキ	名詞-普通名詞-形状詞可能	*	*	セイシキ	4	-1	-1
4	に	ニ	だ	ダ	助動詞	助動詞-ダ	連用形-ニ	ニ／	1	*	-1
5											
6	決まっ	キマッ	決まる	キマル	動詞-一般	五段-ラ行-一般	連用形-促音便	キマッ	3	-1	-1
7	た	タ	た	タ	助動詞	助動詞-タ	基本形-一般	タ／	1	*	-1
8											
9	わけ	ワケ	訳	ワケ	名詞-普通名詞-一般	*	*	ワ'ケ	2	1	1
10	で	デ	で	デ	助詞-格助詞	*	*	デ'	1	*	1
11	は	ハ	は	ハ	助詞-係助詞	*	*	ハ／	1	*	-1
12											
13	ない	ナイ	無い	ナイ	形容詞-非自立可能	形容詞-ア段-無イ+ない	基本形-一般	ナ'イ	2	1	1
14	の	ノ	の	ノ	助詞-準体助詞	*	*	ノ'	1	*	1
15	で	デ	だ	ダ	助動詞	助動詞-ダ	連用形-一般	デ	1	*	-1
16											
17	カネ	カネ	金	カネ	名詞-普通名詞-一般	*	*	カネ	2	-1	-1
18	の	ノ	の	ノ	助詞-格助詞	*	*	ノ	1	*	-1
19	カ	チカラ	カ	チカラ	名詞-普通名詞-一般	*	*	チカラ'	3	3	3
20	も	モ	も	モ	助詞-係助詞	*	*	モ／	1	*	-1
21											
22	十分	ジュウブン	十分	ジュウブン	副詞	*	*	ジュウブ'ン／	4	3	3
23											
24	知っ	シッ	知る	シル	動詞-一般	五段-ラ行-一般	連用形-促音便	シッ	2	-1	-1
25	て	テ	て	テ	助詞-接続助詞	*	*	テ	1	*	-1
26	い	イ	居る	イル	動詞-非自立可能	上一段-ア行	連用形-一般	イ	1	-1	-1
27	た	タ	た	タ	助動詞	助動詞-タ	基本形-一般	タ	1	*	-1
28											
29	首都	シュト	首都	シュト	名詞-普通名詞-一般	*	*	シュ'ト／	2	1	1
30											
31	バンコク	バンコク	バンコク	バンコク	名詞-固有名詞-地名-一般	*	*	バ'ンコク	4	1	1
32	で	デ	で	デ	助詞-格助詞	*	*	デ'	1	*	1
33	は	ハ	は	ハ	助詞-係助詞	*	*	ハ／	1	*	-1

Fig. 1 形態素解析結果と対応付けたアクセントラベル

実際に学習と推定を行なうにあたり、データベースに含まれるアクセント句境界位置情報を使用して文をアクセント句に区切り、句に含まれる各形態素の文中アクセント型を学習と推定の対象とした。ここでいう文中アクセント型とは、「音声合成」であれば「オンセー」（無核）と「ゴーセー」（1 モーラ目に核）を指す。すなわち、「音声（オンセー）」と「合成（ゴーセー）」から「音声合成（オンセーゴーセー）」が作られるアクセント結合現象を、形態素「音声」「合成」のそれぞれのアクセント型が変化した（単独発声アクセント型とは異なる文中アクセント型になった）ものと見做して扱う。文中で形態素の境界にアクセント核が生じているもの（「流れ・を（ナガレ・オ）」「出来・ない（デキ・ナイ）」など）では、当該アクセント核は前側の形態素に属するものとした。このようにして形態素解析結果、単独発声/文中アクセント型をまとめた結果が Fig.1 であり、列 H が文中アクセントラベル済みテキストを形態素に対応付けたもの、列 K が形態素の文中アクセント型、列 J が形態素の単独発声アクセント型である。なお、“−1” はアクセント核が存在しないことを表し、“*” は単独アクセント型ラベリングの対象となっていない形態素であることを表す。その他の数値は核のあるモーラの位置を表している。

この過程で、形態素内にアクセント句境界が存在している文、複数のアクセント核を持つ形態素を含む文は文ごと除去した。

これらの文を学習用 3,581 文（25,692 アクセント句）、推定用 527 文（3,533 アクセント句）に分けて使用した。この区分けは、データベースの基となる JNAS 新聞記事文の 35 ファイルと 5 ファイルとして分割したものである。

2.3 単独発声アクセント型を与えない場合における文中アクセント型の学習・推定

前節のデータを使用して、文中アクセント型の学習と推定を行なった。まず、形態素単独発声アクセント型が与えられていない状況を想定するものとし、

- 観測素性
  前後 2 形態素を含めた 5 形態素について、

*Statistical Learning of Accent Sandhi for Developing Japanese Text-to-speech Systems, by Ryo Kuroiwa , Nobuaki Minematsu , Keikichi Hirose (University of Tokyo)

Table 1　CRF による文中アクセント型推定の結果

	すべての句		単純な句		複合名詞を含む句	
直接学習（単独型あり）	2833 /3533	82.1%	703 /822	85.9%	530 /688	77.0%
直接学習（単独型なし）	3081 /3533	87.2%	775 /822	94.3%	523 /688	76.0%
変化の学習	3137 /3533	88.8%	791 /822	96.2%	553 /688	80.4%

– [基本形/基本形読み/書字形/品詞/活用型]（組み合わせ; 形態素の同定に用いる）
– [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]

と当該形態素の文中アクセント型との関係

- 遷移素性（隣接する形態素の文中アクセント型同士の関係）

を使用した。その結果が、Table 1 の上段である。約 82% のアクセント句について正しい推定が得られている。このうち、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）については正解率が高く、逆に名詞の連続（複合名詞）を含むアクセント句では率が低い様子が見られたため、それぞれ別途に集計してある（2, 3 列目）。なお、複数のアクセント核を持つアクセント句については、1 つめの核（主アクセント）が一致していれば正しい推定ができたものと見做した。すべてのアクセント核が一致した場合のみをアクセント句としての正解とすると、最大 3%程度一致率が低下する。

2.4　単独発声アクセント型を与える場合における文中アクセント型の直接学習・推定

前節での素性に加え、観測素性に

- – [単独発声アクセント型]

を加えて同様の学習を行なった。その結果が、Table 1 の 2 段目である。

2.5　単独発声アクセント型から文中アクセント型への変化の学習・推定

前節の手法では、文中アクセント型を直接学習や推定の対象としていたため、単独発声型と文中アクセント型がいずれも “1” である場合といずれも “2” である場合は別々の事象として扱われることとなる。そこで、単独発声アクセント型から文中アクセント型への変化を学習・推定の対象とすることで、類似する現象を共通のものと捉えられるようにする。具体的には、

- 単独発声アクセント型が有核のもの（“−1” 以外の値を持つもの）に対しては、
  – 文中アクセント型が有核の場合、単独発声型からの変化量を表すラベル（“[0]”, “[+1]”, “[+2]”, ...; “[−1]”, “[−2]”, ...）
  – 文中アクセント型が無核（“−1”）の場合、無核を示すラベル（“non”）
- 単独発声アクセント型が無核のもの（“−1”）または値を持たないものに対しては、
  – 文中アクセント型（“1”, “2”, ...; “−1”）

を機械的な処理によって与え、そのラベルを前節と同様の素性によって学習・推定し、推定結果のラベルを再び機械的に文中アクセント型を表す数値に復元した。その結果が、Table 1 の下段である。

2.6　考察

文中アクセント型を直接学習・推定する場合、単独発声のアクセント型を用いないもの（Table 1 の上段）と用いるもの（2 段目）では多くの違いが見られ、特に単純な結合においては 86%から 94%になるなど顕著な違いが見られる。しかし、複合名詞を含むアクセント句では一致率に向上が見られない。単純な結合では単独発声時のアクセントがそのまま文中アクセントとなることが多いのに対し、複合名詞においては単独発声時アクセントに関わらないアクセント変化が多いことによるものであろう。

アクセントの変化を学習・推定の対象としたもの（下段）では、アクセント型を直接推定するもの（2 段目）と比較し、全体として均一に一致率が向上している。単純な結合でも 4% ほどの不一致が見られるが、不一致の中にも、アクセントの揺れの範囲に納まると考えられるものが多数存在し、推定性能としては十分に高いもの考えられる。それに対し、複合名詞を含むアクセント句では依然 20%程度の不一致が見られる。複合名詞では、文脈や係り受け関係によって結合の仕方に違いが出ることがあり、揺れも多い。例えば、「政権交代」は「セーケンコータイ」「セーケンコータイ」のいずれにもなり得る。また、「ソ連邦解体後」「日米交換船」のようなアクセント句では係り受けの影響を受ける可能性が高い。このような問題は規則に基づくアクセント結合処理 [1] でも解決できていない問題であり、新たな手法が必要であると考えられる。

3　おわりに

統計的手法を用いることで、単純なアクセント結合については完全に近い精度で処理ができ、日本語テキスト音声合成システムへの導入も可能な水準に達するものであるといえる。しかし、複合名詞など複雑な構造を持つものは統計的手法においても幾らかの誤りが発生することは避けられず、新たな改善策が必要となるところであろう。

参考文献

[1] 匂坂, 佐藤:“日本語単語連鎖のアクセント規則”, 電子通信学会論文誌 J66-D.7, pp.847-856, 1983.

[2] 喜多 他: “日本語テキスト音声合成を目的としたアクセント結合規則の構築と改良”, 電子情報通信学会 信学技報，SP2002-26，pp.13-18, 2002.

[3] 長野 他: “確率モデルを用いた読み及びアクセント推定”, 情報処理学会研究報告, Vol.2005 No.69 pp.81-86, 2005.

[4] J. Lafferty *et al.*: “Conditional Random Fields: Probablistic Models for Segmenting and Labeling Sequence Data”, Proceedings of the 18th International Conference on Machine Learning, pp.282-289, 2001.

[5] 工藤 拓: “CRF++: Yet Another CRF toolkit” http://chasen.org/˜taku/software/CRF++/

[6] 黒岩 他: “特定話者による大規模アクセントラベリングとそのデータベース化”, 日本音響学会 2007 年春季講演論文集 1-Q-18, 2007.