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

印南圭祐[†]渡辺美知子[†]峯松信明[†]広瀬啓吉[‡][†]東大・新領域[‡]東大・情報理工

1 はじめに

日本語テキスト音声合成システムにおいて日本語文の自然な読み上げを実現する場合、文中のアクセント核の位置を適切に推定する必要がある。日本語単語のアクセント核位置は、孤立発声時と、複数の語が接続した文中の発声時とでは容易に変化する(アクセント結合)。よって、辞書に登録されている孤立発声時のアクセント核位置情報や言語情報を参照して、文中のアクセント核位置を適切に推定するモジュールの開発が求められる。従来は、規則に従いアクセント結合を処理する手法 [1] が用いられてきた。一方筆者等は、CRF(条件付き確率場)を用いて、単語の孤立発声時のアクセント核位置から文中のアクセント核位置を統計的に推定する手法を提案し、規則による手法よりも高精度なアクセント結合の処理を実現した [2][3]。この手法は、学習データの増加や素性の再設計によって精度向上が見込める。本報では、CRFに基づく処理モデルに関して、1) 学習データ増加による推定精度の向上、2) 誤推定結果の分析とそれに基づく改良、3) アクセント結合推定の前処理として必要となるアクセント句境界推定、を検討したので報告する。

2 CRF を用いた統計的アクセント結合処理の先行研究

筆者等はまず、JNAS[4] 及び ATR 音素バランス 503 文のテキストに対し、単独ラベラによるアクセント句境界及び核の位置情報を登録したデータベースを作成した [2][3]。ラベリング内容がラベラの方言・アクセント感覚に依存するため、全てのラベリングは単独ラベラによって行われた。その上で、ラベリングが終了した読み上げ文 40 セット (4,108 文) を、学習用/評価用として 35 セット (3,581 文) / 5 セット (527 文) に分割し、CRF++[5] を用いて文中の各形態素のアクセント核位置を推定する実験を行なった。結果の集計は形態

素単位とアクセント句単位で行なった。特に後者では、主アクセントのみの正誤判定と、副次アクセントを含めた全てのアクセントに着目した判定を行なった。なお、アクセント句中で最初に出現するアクセント核を主アクセントとし、それ以降の形態素に与えられたアクセント核は全て副次アクセントであるとみなしている。更にエラー解析として、下記の 2 種のアクセント句に着目した。

- 単純なアクセント句
{ 自立語 } + { 付属語 } から構成され、アクセント結合が 1 回だけ起こるアクセント句
- 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

上記 2 種のアクセント句の推定精度を参考にして素性の再設計を行ない、推定精度の向上を試みた。その結果をまとめて、規則を用いた手法と共に表. 1 に示す。素性の 3 段階の改良内容については、以下に示す。素性の詳細情報は [2][3] を参照して戴きたい。

1. 単独発声アクセント型からの変化量の学習・推定

文中のアクセント核位置を直接推定するのではなく、孤立発声時のアクセント核の位置からの変化量を示す相対変化ラベルを学習・推定対象とすることで、直接推定する場合より全体的に処理精度が向上した。

2. 隣接形態素組み合わせ素性を用いた学習・推定

推定対象となる形態素とその前後の形態素のラベルの組み合わせ素性を学習に利用することで、特に名詞連続を含むアクセント句の処理精度が向上した。

3. アクセント結合規則に適合させた学習・推定

表 1: 各手法のアクセント結合処理精度の比較

	形態素	すべての核			主核のみ		
		すべての句	単純	名詞連続	すべての句	単純	名詞連続
規則に基づく手法	—%	76.4%	94.4%	73.5%	76.8%	94.5%	74.2%
CRF を用いた手法	93.6%	85.0%	93.6%	74.3%	87.2%	94.3%	76.0%
+相対変化ラベル	94.1%	86.3%	95.9%	78.5%	88.8%	96.2%	80.4%
+組み合わせ素性	95.1%	88.6%	95.5%	82.3%	91.0%	96.1%	84.0%
+規則に基づく変化ラベル	95.5%	89.4%	95.5%	83.4%	91.7%	96.4%	85.6%
学習量を増加させた CRF	96.5%	91.9%	97.2%	86.6%	93.5%	97.7%	87.9%

表 2: 学習データ量の変化に伴う誤推定の減少率 (1 セット当たり約 100 文)

セット数	形態素	すべての句	単純	名詞連続
10 → 20	18.5%	15.6%	9.8%	15.5%
20 → 30	21.0%	21.1%	15.2%	20.3%
30 → 40	9.5%	10.0%	7.7%	13.1%
40 → 50	10.5%	11.1%	0.0%	9.4%
50 → 60	10.8%	10.9%	25.0%	6.3%
60 → 65	3.9%	2.7%	14.8%	-2.2%

句坂らのアクセント結合規則 [1] には出現しやすい結合のパターンがあり、これらを特に考慮した相対変化ラベルを学習・推定対象とした。また、規則処理に関係するラベルを観測素性に加えた。その結果、全体的に最も高い処理精度を示した。

3 アクセントデータベースの拡張

3.1 CRF 学習のデータ量増加

JNAS 読み上げ文が全体で 165 セットあるのに対し、先行研究の実験では 40 セットのみが利用可能であった。そこで、同じ単独ラベラによるラベリング作業を引き続き行なうことで、先行研究で作成したデータベースを拡張した。2008 年 1 月の時点で、利用可能な読み上げ文を 70 セット (約 7,200 文) まで増加させた。これを学習/評価用として 65 セット (約 6,500 文) / 5 セット (約 500 文) に分割した。なお、先行研究の評価実験との比較のため、評価データは先行研究のものと同じの文セットとし、学習データのみを 30 セット新たに追加している。これらのデータを用いて、アクセント核位置の推定実験を行った。結果を表 1 最下段に示す。この結果から分かるように、学習データ量を増加することで、CRF によるアクセント結合処理精度は全体的に向上した。

3.2 学習データ量の段階的变化

学習データ量の増加による推定精度の向上の過程と、その飽和性を確認するため、学習データを分割し、段階的に変化させた場合の推定精度を調べた。評価データは引き続き先行研究のものと同じとし、学習データとして用意された 65 セットを 10 セット (約 1,000 文) ずつ増加させ、アクセント句単位の誤推定の減少率を調べた。また、合計 60 セットから 65 セットまで増加させた場合も、10 セットの増加と同様に集計した。なお、アクセント句中の副次アクセントを含めた全てのアクセントに着目して正誤判定を行なった。結果を表 2 に示す。全体的に学習データの総量の増加に伴い、誤推定の減少率は降下する傾向にある (推定精度の飽和)。また、60 セットから 65 セットに増加した場合、単純なアクセント句を除き、減少率が非常に低い。特に、推定精度が伸び悩んでいる名詞連続を含むアクセント句においては、学習データ量の増加により逆に誤推定が増加している (単純な句の減少率の傾向が異なる点については、現在理由を検討中である)。

以上をふまえ、今後 CRF によるアクセント結合処理を改善させるには、学習データ量の増加より、誤り分析に基づく素性の改良、及び推定モデルの改良がより重要になると考えられる。

4 CRF 出力ラベルのエラー解析

読み上げ文 65 セットを学習データとした CRF による推定の結果、文中のアクセント核位置が誤推定された形態素は 342 あり、それを 1 つでも含むようなアクセント句は 285 あった。誤推定された 342 の形態素を単独の音声研究者 (第 2 著者) に判定させた結果、完全に CRF の誤推定と考えられるものは 205 (59.9%) あった。また、単独ラベラのラベリング内容とは異なるものの、許容できる範囲の誤りであると判断されたものは 91 (26.6%) あった。このうち、副次アクセントのものは 31 含まれる。そして残りの 46 (13.5%) は、ラベラ

表 3: アクセント句の構成内容による誤推定数・正答率の比較

	すべての句	単純な句	名詞連続	数詞	カタカナ	副助詞	助詞・助動詞連続
該当アクセント句数	3533	822	688	421	347	128	430
誤答数	285	23	92	36	40	20	62
正答率	91.9%	97.2%	86.6%	91.4%	88.5%	84.4%	85.6%

のラベリング内容より、CRF の推定結果の方が自然であると判断された。以上より、今後 CRF の処理モデルの出力について、許容度を正しく考慮することが重要である。他にも {(人名) + さん} や, {~+もの}, {お+~} などの頻度の高い表現において、誤推定が多いことも大きな課題と考えられる。

また、CRF の完全な誤推定のうち、漢字 1 文字 2 モーラ以下の語を含む複合語における誤推定が特に目立った (63/205)。これより、複合語、および接頭辞・接尾辞を含む表現への対策が必要と分かる。

先行研究ではエラー解析にあたり、2 種類のアクセント句の構成 (単純な句・名詞連続句) に着目した。本稿でも、モデルの改良を目的として、アクセント句の構成に着目した誤り解析を行ない、各句構成における精度を検証した。結果を表. 3 にまとめて示す。

4.1 数詞を含むアクセント句

数詞を含むアクセント句は、全体の推定精度が 91.4% であり、すべてのアクセント句 (91.9%) と比較して特別精度が低いわけではない。しかし数詞は「名詞-数詞」というように、名詞のサブカテゴリとして扱われる。よって、表. 4 のように名詞連続を含む句と重複するものが多く、そのような句の正答率は大幅に低下する (85.6%)。また、形態素解析上、数字表現を桁ごとに別の形態素として扱っている。これにより、桁が大きい数値は数詞が連続で出現しやすく、これも名詞連続と重複する。従って数詞が連続出現するアクセント句への対策が有効だと考える。

4.2 カタカナ表現を含むアクセント句

カタカナ表現の語は外来語、特に名詞が大半を占める。ただし、表. 3 及び 4 から分かる通り、名詞連続のカテゴリと重複する場合と、そうでない場合との精度は殆ど変わらない。よって、カタカナ表現は、名詞との接続時に生じる誤推定より、カタカナ表現自体に起因する誤り解析を行なうべきであると考えられる。

表 4: 評価データ中の分析カテゴリの重複

	名詞連続	数詞&名詞連続	カタカナ&名詞連続
該当句数	688	132	120
誤答数	92	19	14
正答率	86.6%	85.6%	88.3%

4.3 副助詞を含むアクセント句

副助詞は、文中に出現する頻度は高くない。しかし、これを含むアクセント句の推定精度は非常に低い。更に、誤推定された 285 のアクセント句中に出現する副助詞の形態素は合計 21 あったが、この内正しく処理されたものは 3 つしかなかった。今後、副助詞とその前後の表現について更に分析を行なう必要があると考える。

4.4 助詞・助動詞連続を含むアクセント句

高精度の処理が実現できるような単純なアクセント句では、助詞・助動詞は連続で出現しない。そのような助詞または助動詞が連続して出現するアクセント句の精度を確認したところ、これまで定めた分析カテゴリ中、副助詞を含む句に次いで低かった。これと副助詞を含むアクセント句をふまえ、助詞・助動詞を含むアクセント結合処理に対する改善が必要と言える。

4.5 エラー解析に基づく改良の検討

エラー解析より、名詞連続、及び付属語への対策が必要といえる。利用しているアクセントデータベース中の付属語は、自立語と違って孤立発声時のアクセント型の情報が登録されていない。そこでアクセント辞書の記述に従って付属語に対し、孤立発声時のアクセント型を含む、アクセント結合に関する属性ラベルの追加を検討している。現在新たな属性として、付属語がアクセント結合後にアクセント核を持ちうるかどうかを示すラベルの導入作業を行なっている。

表 5: CRF によるアクセント句境界位置の推定結果

ラベラの指定境界	CRF の推定境界	CRF とラベラの境界一致	脱落誤り	挿入誤り	Recall	Precision
8,329	7,839	7,561	768	278	90.8%	96.5%

5 CRF を用いたアクセント句境界位置の推定処理

CRF を用いたアクセント変形の推定処理は、入力文のアクセント句境界位置が既知であることを前提としている。しかし、日本語テキスト音声合成システムのモジュールとしての利用を想定する場合、未知の文章に対してもアクセント句境界位置が適切に決定される必要がある。そこで入力文の形態素解析結果から、各形態素の直前にアクセント句境界が存在するかどうかを、CRF を用いて統計的に推定する手法を検討した。

5.1 評価実験

学習・評価データには、アクセント結合処理の実験と同じアクセントデータベースを利用できる。ただし、JNAS の読み上げ文に含まれる句読点、および一部の補助記号の情報が取り除かれているため、ラベリング対象となったテキスト原文に含まれている句読点等の情報をアクセントデータベースに追加した。

こうして利用可能にした読み上げ文 70 セットを、学習/評価データとして先頭から 60 セット/末尾 10 セットに分割した。CRF の観測素性には、「出現形」「品詞の詳細情報」の 2 つを利用した。また、「句読点の直後に必ずアクセント句境界が存在する」という特徴の学習を考慮し、素性には推定対象の形態素と前後 1 形態素の特徴を含めた。実験結果を脱落誤り・挿入誤りを含めて表 5 に示す。単純な構成の素性で、Recall で 90.8%、Precision で 96.5% の結果が得られた。

5.2 エラー解析

推定結果から脱落誤りと挿入誤りをそれぞれ 50 箇所抽出し、単独の音声研究者(第 2 著者)が、許容できる誤推定かどうかの判定と境界前後の品詞に基づくエラー解析を行なった。その結果、許容できない誤りは脱落誤りの場合 25/50(50.0%)、挿入誤りの場合 37/50(74.0%)あり、挿入誤りの方がエラー率が高いと分かった。許容できない誤りの前後の品詞を調べた結果、名詞連続に対する誤推定が特に多く、脱落誤りでは 10/25(40.0%)、挿入誤りでは 17/37(45.9%)であると分かった。また、

許容できない挿入誤りでは、先行する語が助詞である場合が 9/37、接辞である場合が 4/37 と多く、助詞では特に「の」が先行して出現するケースが多く、5/9 であった。今後は、素性の再設計によるエラーの減少を検討していく。

6 まとめ

CRF によるアクセント結合処理は、学習データの増加によって改善させられるが、その向上には限界がある。そこでエラー分析の結果から、名詞連続を含む句と、助詞・助動詞等の付属語に対するアプローチが重要だと考える。今後は誤り分析の強化と、それに基づくデータベース・CRF の学習素性の改良を行ない、評価実験にて検証する予定である。また、アクセント結合処理の前処理として、アクセント句境界の推定を CRF で行なう手法も検討した。その改善には、人の判断による推定結果の妥当性の検証と、素性の再設計によるモデルの改良が重要になる。

謝辞 本研究に全面的にご協力頂いた黒岩龍様と、度々御助言を頂いた特定領域研究「日本語コーパス」電子化辞書班の皆様へ厚く感謝申し上げます。

参考文献

- [1] 句坂, 佐藤, 信学論, J66-D7, pp.849-856, 1983.
- [2] 黒岩他, 信学技報, SP2006-174, pp.31-36, 2007.
- [3] 黒岩, “日本語音声合成のためのアクセント結合規則の改善とデータベースに基づく統計的アクセント処理”, 東京大学大学院情報理工学系研究科電子情報学専攻修士論文 2007.
- [4] <http://www.mibel.cs.tsukuba.ac.jp/jnas/>
- [5] <http://crfpp.sourceforge.net/>