ローマ字表記単語の読み上げのための長音置換・追加位置判定法

浅野久子　中嶋秀治　水野秀之　奥雅博

日本電信電話株式会社　ＮＴＴサイバースペース研究所

1. はじめに

日本語テキストにはアルファベット単語が数多く含まれる場合がある。例えば、あるWebサイトの約8,000店の「買う」というジャンルの店舗情報には、延べで約6,400語、異なりで約2,900語のアルファベット単語が存在する。これらのアルファベット単語は店名などの固有名詞が多く、すべてを辞書登録するのは非現実的であるため、テキスト音声合成(TTS)で正しく読み上げるためには、アルファベット未知語に対し、カナの読みを付与する必要がある。

これらのアルファベット単語は様々な言語由来の語が含まれるため、我々は、アルファベット未知語に対して、ステップ1：英語読み、アルファベット読み、ローマ字読みといった読みクラスに分類し[1]、ステップ2：読みクラス別の読み付与処理を行うという2段階のアプローチをとっている。本稿では、このうち、ステップ2の中のローマ字読み付与処理の精度を向上させる新手法を提案する。

ローマ字単語[1]に対する読み付与は、基本的には、従来から行われている対応表ベースのローマ字カナ変換の手法、すなわち、ローマ字ーカナ対応表（例：'ta'→タ）、および、促音・撥音に対する特殊ルール（子音が連続する部分が促音など）を用いればよく、我々もこの手法を用いてきた。しかし、TTS用の読み付与で用いる場合、長母音、すなわち読みの長音化に対応できないという問題があった。長母音は、発声が引き伸ばされる母音である。例えば、「雄三」はふりがなとしては「ユウゾウ」と書くが、実際の発声としては、「ユーゾー」と2つの「ウ」の部分を引き伸ばして発声する（「ユ」「ウ」と分けて発声しない）。ローマ字表記においては、'aa', 'ii', 'uu', 'ee', 'oo', 'ei', 'ou'の7種類（以後、これらを長音置換候補文字列とよぶ）が長母音である可能性がある[2]。

ローマ字の読み付与における長音化の課題は2種類ある。一つはローマ字表記の特徴ともいえる、長母音の表記上の省略である。前述の7種の長母音のうち'ei'以外のものは、2番目の母音を省略して書く場合がある（「雄三」→'Yuzo'）が、対応表による読み付与ではこれに対応できない（「ユゾ」）。正しい読みとするためには、適切な位置に長音を追加する必要がある（「ユーゾー」）。

もう一つの課題は、長母音をローマ字表記上省略しなくても、長音化には対応できないというものであり、TTS特有の課題といえる。例えば、'Yuuzou'は、対応表による読み付与では「ユウゾウ」となる。人間の発声では、「ユーゾー」と長音化する代わりに、区切って「ユウゾウ」と発声してもさほど不自然ではない。しかし、TTSにおいては、特にコーパスベースの合成方式[2]を用いる場合、音声データベース上「ユー」と「ユウ」は別々の音としてラベル付けされ全く異なる音として扱われるため、通常長音化すべきところを長音化していないと、致命的な音質の劣化となる場合がある。そこで、対応表によりカナ変換された長母音（「ユウゾウ」）を長音に置換する（「ユーゾー」）必要がある。

我々は、これら2つの長音化の課題に対して、長音追加・長音置換が生じる可能性のある最小単位、すなわちモーラ（ほぼカナ1文字に相当）単位に、綴りの情報のみを用いて、長音追加・長音置換・そのままの3値いずれかに分類する長音化分類問題として扱う統計的アプローチをとる。本稿では、分類器として、Support Vector Machine（SVM）[3]をpairwise法などにより多値分類に拡張した汎用TaggerであるYamCha[3]を使用した。SVMを用いたのは、近年、様々な自然言語処理タスクに適用され、その有効性が報告されているためである。

2. 読み付与処理の流れ

ローマ字単語に対する読み付与処理の流れを図1に示す。はじめにローマ字ーカナ対応表を用いてカナに変換し、モーラ単位に分割する。ここで、読みの多義が生じる'n'（撥音（ン）か、ナ・ニャ行か）については、頻度の高い以下の読みを付与する。

- 'n'+子音：撥音（ン）。ただし行頭で'n'+'y'の場合のみニャ行とする。例：ko***n***yaコ***ン***ヤ
- 'n'+母音：ナ行。例：e***na***tsuエ***ナ***ツ

[1] 本稿では、ローマ字単語を、アルファベット表記された日本語単語と定義する。外来語のローマ字表記（例：Rondon＝ロンドン）は考慮しない。

[2] 長母音にならないのは、漢字表記した場合の文字境界（＝言葉の意味をなす最小単位）を跨る場合である。例：Tat***ei***shi（立石）→○タ**テイ**シ、×タ**テー**シ

[3] http://cl.aist-nara.ac.jp/~taku-ku/software/yamcha/

次に、各モーラ単位に長音化分類を行い、分類結果を反映させて、最終的に読みを出力する。

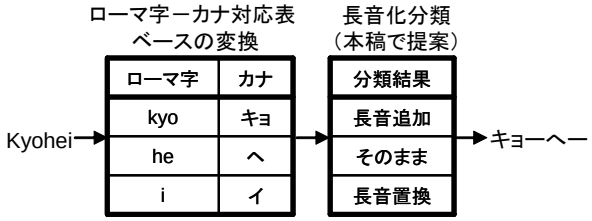

図 1 ローマ字単語に対する読み付与処理の流れ

3. SVM による長音化分類

SVM で利用されるカーネル関数としては様々なものが提案されているが、本稿では YamCha に組み込まれている d 次の多項式

$$\mathrm{K}(\mathbf{x},\mathbf{y}) = (\mathbf{x}\cdot\mathbf{y}+1)^d \qquad (1)$$

を利用する。d 次の多項式は d 個までの属性の組み合わせを考慮した学習モデルと見なすことができる。最適な d は 6.4 節で検証する。

長音化分類では、ローマ字単語の各モーラを長音追加・長音置換・そのままの 3 種類に分類する。SVM は 2 値分類器であるため、多値分類のための拡張を行う必要がある。本稿ではこの拡張として pairwise 法、すなわち、各 2 クラスを判別する 3 種類の分類器を作成し、それらの多数決で最終的にクラスを決定する手法を採用する。

YamCha は、属性として静的属性と動的属性の 2 種類を扱える。静的属性は、分類を行う前にその値が決定できる属性であり、長音化分類においては、綴りから得られる以下の 3 属性を扱う。

- モーラ表記：モーラの表記そのもの。
- 行：モーラの行情報。ヘボン式・訓令式のゆれは吸収（正規化）する。
- 段：モーラの段情報。

例えば、'ta'（タ）というモーラのモーラ表記='ta'、行='t'、段='a'となる。また、'chi', 'ti'の行はいずれも't'と正規化する。撥音（ン）と促音（ッ）については、行、段とも専用の値（撥音='N'、促音='T'）を割り当てる。長音化分類に最適な綴り情報の組み合わせについての検証は 6.2 節で行う。

動的属性は、分類時に順次得られる分類結果であり、長音化分類では前方のモーラの分類結果となる。

あるモーラ（当該モーラ）の長音化分類をする際に属性として利用するモーラの範囲（文脈）は、当該モーラとその前後 n モーラ（動的属性は前 n モーラ）とする。最適な文脈長 n は 6.3 節で検証する。

4. 実験用データ

ローマ字単語は、固有名詞の占める割合が非常に高いと考えられる。そこで、TTSで利用している単語辞書（長音化情報付きの読みを格納）から、表記が漢字とひらがなのみからなる姓名 26,523 語（読み異なり語）を抽出し、機械的にカナローマ字変換した。この変換では、すべてのゆれを全展開した。すなわち、'ei'以外の長音置換候補文字列は、2 番目の母音を省略するものとしないもの（例：オーヤマ→oyama, ooyama[4]）の 2 種類に展開し、ローマ字表記法のヘボン式と訓令式の差（例：チ→'chi', 'ti'）も展開した。この結果、49, 031 語のローマ字単語が得られた（姓名ローマ字データ）。

また、実際に出現するローマ字単語データとして、Web 上の約 8,000 店の店舗情報から、254 語の異なりローマ字単語も抽出した（Web ローマ字データ）。

5. ベースラインと評価基準

5.1. 長音置換

長音置換におけるベースラインは、長音置換候補文字列（'aa', 'ei'等 7 種）の 2 文字目に相当するカナを長音に置換するかどうかを、その表記別の多数派に一意に決定することとする。4 節の姓名ローマ字データで検証したところ、'aa', 'ee'は長音置換しない、'ii', 'uu', 'oo', 'ei', 'ou'は長音置換するが多数派であったため、そのように読みを付与する。この場合、'aa', 'ee'はこれが漢字表記上の文字境界に跨らない場合、'ii', 'uu', 'oo', 'ei', 'ou'は漢字表記上の文字境界に跨る場合が誤りとなる。

長音置換の評価は、長音置換候補文字列の 2 番目にあたるモーラを対象に、

$$\text{長音置換正解率} = \frac{\text{長音置換判定が正しい数}}{\text{長音置換候補文字列総数}}(\%) \qquad (2)$$

を用いる。

5.2. 長音追加

長音追加におけるベースラインは、長音追加を全く行わなかった場合とする。

長音追加の評価は、長音置換候補文字列以外の母音を含むモーラを対象に、

$$\text{長音追加正解率} = \frac{\text{長音追加判定が正しい数}}{\text{母音を含むモーラ総数}}(\%) \qquad (3)$$

を用いる。ここで、(3)式の分母の母音としては、(2)式の分母となる母音は除く。

5.3. 単語正解率

総合的な評価尺度として、

$$\text{単語正解率} = \frac{\text{読みが完全に正しい単語数}}{\text{総単語数}}(\%) \qquad (4)$$

を用いる。

[4] 'ohyama'という変換もありうるが、'oh'という表記は'オー'と長音化すると対応表ベースのローマ字カナ変換で一意に判定できるため、今回のデータには含めない。

6. 実験

4 節で示した姓名ローマ字データを用いて、各種パラメータや学習データ量を変動させ、最適な設定の検証実験を行った。この実験では、姓名ローマ字データ 49, 031 語を 10 等分してクロスバリデーションを行い、(2)～(4)式で定義した各正解率の平均正解率で評価した。また、最適設定を対象に誤りの分析を行い、さらに、実データである Web ローマ字データでの評価も行った。

6.1. ベースライン

はじめに、5 節で示したベースライン法の、クロスバリデーションの各テストデータに対する平均正解率を表 1 に示す。さらに、参考として平均対象数、すなわち(2)～(4)式の分母の平均値も示す。長音追加の方が長音置換より圧倒的に対象数が多い。

表1 ベースラインの平均正解率・平均対象数

	長音置換	長音追加	単語
平均正解率	89.3%	94.4%	80.3%
平均対象数	1,191	16,388	4,903

6.2. 綴り情報

有効な綴り情報を検証するために、

- モーラ表記、行、段
- モーラ表記
- 行、段

の 3 種類の静的属性セットを用いた結果を表 2 に示す。その他のパラメータは、文脈長 n=2、カーネル関数の次数 d=2 とした。

表2 属性セット別の平均正解率

属性セット	長音置換	長音追加	単語
モーラ表記,行,段	98.2%	97.7%	91.9%
モーラ表記	98.2%	98.0%	92.7%
行,段	98.2%	98.2%	93.2%

長音追加に対しては、行と段の 2 属性を使うものが最もよい。これは、モーラより行・段に情報を分割した方が、データのカバー率が上がりロバストになること、および、ア段はほとんど長音追加しないがオ段はよくするなど、段情報が重要となるからだと考えられる。

これに対し、長音置換で属性セットによる違いがないのは、長音置換は母音が連続表記された部分で例外的に長音置換されない条件をつかむのが重要であり、モーラ表記でもこの条件を十分カバーできるからであろう。

6.3. 文脈長

文脈の有効性を検証するため、6.2 節で最も有効であった属性セット「行、段」に対して、文脈長 n=1,2,3 とした結果を表 3 に示す。その他のパラメータは 6.2 節と同一である。

表3 文脈長別の平均正解率

文脈長(n)	長音置換	長音追加	単語
1	92.6%	96.7%	88.1%
2	98.2%	98.2%	93.2%
3	95.5%	97.5%	91.0%

文脈として前後 2 モーラを使ったものが最もよい。漢字 1 文字のモーラ数は 2 以下であるものが大多数であることから、長音化分類における文脈は、前後の漢字 1 文字相当で十分であるといえる。

6.4. カーネル関数の次数

(1)式のカーネル関数における最適な次数 dを検証するため、6.2 節で最も有効であった属性セット「行、段」に対して、d=1,2,3 とした結果を表 4 に示す。その他のパラメータは 6.2 節と同一である。

表4 次数 d 別の平均正解率

次数(d)	長音置換	長音追加	単語
1	98.3%	98.1%	93.1%
2	98.2%	98.2%	93.2%
3	98.3%	98.1%	93.0%

長音化分類においては、d による差はほとんどないといえる。

6.5. 学習データ量

適切な学習データの量を検証するために、クロスバリデーション用に 10 等分したデータの 1 つを共通の評価データとし、残りの 9 つのデータのうち 1 つ～9 つ（9 つで 44,128 語）を学習データとして用いた場合の正解率を図 2 に示す。

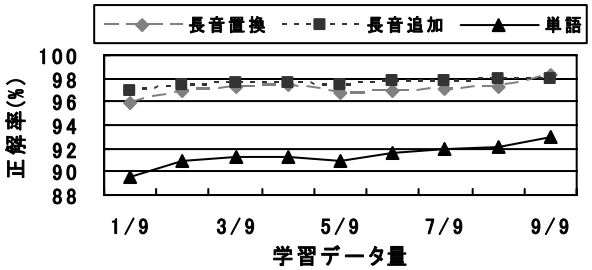

図 2 学習データ量別の正解率

長音置換・長音追加とも、学習データ量が 1/9 では、正解率の低下の度合いが他に比べて大きく、不十分な量といえる。それからデータ量を増やしていくと、5/9 までは正解率に増減がみられるが、その後 6/9～9/9 まで微増していく。これより、約 1 万語（2/9 の量）の学習データを用いれば、ある程度十分な精度が得られ、今回用意した全て（約 4 万 4 千語）

を用いれば、データ量増加に対する効果の度合いは低くなるものの最もよい結果が得られるといえる。

6.6. 最適設定

以上より、今回の実験において、長音化分類に対する最も有効な設定は、

- 綴り情報（静的属性）：行、段
- 文脈長：n=2
- カーネル関数の次数：d=2
- 学習データ量：約 44,000 語

であり、平均単語正解率 93.2%を達成した。ベースラインの平均単語正解率は 80.3%であり、本手法は、単語誤りを約 65%減らすことができた。

6.6.1. 誤りの分析

上記の最適設定において、クロスバリデーションの 1 セットにおける誤りの分析を行った。誤り単語 349 語の単語誤りの分布を表 5 に示す。ここではローマ字単語の読みのあいまい性に着目して分類した。ローマ字単語では長音にあたる表記が省略されることがあるが、その長音省略された表記と、元々長音がない語の表記が同一になる場合がある。例えば、'*o*yama'という表記は、「**オー**ヤマ（大山）」か「**オ**ヤマ（小山）」か、表記からだけでは判断できない。そこで誤りとなった語において、付与された読みが姓名として存在しうると判断されるものはあいまい性ありと分類した[5]。表 5 の'n'は、2 節で示した'n'の読み付与が誤っていたことを示す。

表5 単語誤りの分布（語数）

あいまい性	長音置換	長音追加	'n'
あり	2.9% (10)	70.8% (247)	0.6% (2)
なし	2.0% (7)	18.0% (63)	5.7% (20)

全誤りの 7 割以上が単語の読みにあいまい性がある誤りであった。単語の読みにあいまい性がない誤りにおいても、誤った読みと、文脈内（当該モーラとその前後 2 モーラの 5 モーラ）で 3 モーラ以上の情報が一致する学習データをもつ語が[6]、長音置換で 3 語、長音追加で 51 語存在した。これらより、単語としての読みのあいまい性がない語でも、その部分文字列における読みのあいまい性が誤りの主因であるといえ、綴り情報のみからの読み付与としてはほぼ限界に近い精度を達成しているといえる。

長音以外の読み誤りとして'n'の読みのあいまい性があるが、長音に比べると誤り頻度が少なく、誤りも特定の事例に集中している（'Junichiro'など、'イチ'（一）を'ニチ'にしてしまうものが半数以上）ので、'nichi'は'ンイチ'とするといった例外的なルールを作成することで対応できると考えられる。

6.6.2. Web ローマ字データでの評価

4 節で述べた Web ローマ字データ 254 語に対して読み付与を行った結果を表 6 に示す。ここで、提案手法では 6.6.1 節と同じ学習モデルを用いた。

表6 Web ローマ字データでの正解率(正解数/総数)

分類法	長音置換	長音追加	単語
ベースライン	87.5% (14/16)	94.1% (657/698)	87.4% (222/254)
提案手法 (SVM)	87.5% (14/16)	96.4% (673/698)	92.5% (235/254)

提案手法は、Webローマ字データに対しても姓名ローマ字データとほぼ同等の精度を達成し、ベースラインよりもよい結果となっている。このデータは長音化分類対象が少なく[7]、特に長音置換ではベースラインとSVMに差は現れなかったので、厳密な評価とはいえないが、姓名ローマ字データから作成された学習モデルは、固有名詞を含む割合が高いWeb上のデータに対しても有効であると考えられる。

7. おわりに

本稿では、ローマ字単語に対する長音化分類手法を提案し、姓名データに対して、ベースラインに比べて単語の読み誤りを約 65%減らせることを示した。残る誤りの 90%は単語、もしくはその部分文字列の読みのあいまい性が原因であり、本手法は綴り情報のみからの読み付与としては十分な精度を達成していることを確認した。

参考文献

[1] 浅野, 永田, 阿部: 日本語テキストにおけるアルファベット文字列の読みクラス分類, 言語処理学会第 9 回年次大会,2003

[2] A. Black, N. Campbell: Optimising selection of units from speech databases for concatenative synthesis, Eurospeech95, 1995

[3] V. Vapnik: The Nature of Statistical Learning Theory, Springer, 1995

[5] 姓名ローマ字データ内であいまい性があるもの、および、姓名ローマ字データ内ではあいまい性がなくても、姓名として存在しうると人手で判断したものをあいまい性ありとした。

[6] 例えば、'*o*kusu'=「**オー**クス」の先頭モーラ'o'の判定を「そのまま」と誤った。学習データ中には、'*o*kusumi'=「**オ**クスミ」が存在した（文脈内'[先頭]okusu'完全一致）。

[7] Webローマ字データにおいて、長音置換が必要な単語は 12 語（うち 7 語は省略表記不可能な'ei'）、長音追加が必要な単語は 30 語であった。これより、ローマ字表記上、長音は省略される割合の方が高く、長音置換より長音追加がより重要であると考えられる。