RE-002

文単位で分割されたテキストで学習した言語モデルによる単語信頼度を用いた文境界検出
Sentence Boundary Detection Using Confidence Measure Based on Language Model Trained with Sentence

鈴木 伸尚† 西田 昌史† 山本 誠一†
Nobuhisa Suzuki Masafumi Nishida Seiichi Yamamoto

1．はじめに

近年，音声認識による音声要約や会議の自動書き起こしなどの研究が進められている．従来の音声認識では，一定のポーズで発話を分割して認識を行い，その結果を出力するのが一般的である．しかし，ポーズは書き言葉でいう「節」や「文」とは無関係に出現することがしばしばある．特に，講演音声に対してはポーズ長だけで文境界を判断すると，閾値を超えるポーズでも文境界とならない部分が多いとされている[1]．その結果，例えば議事録の自動書き起こしならば，意味的なまとまりで区切れることなく認識結果が出力されてしまい，読みづらく内容を理解するのが困難となってしまう．

このような問題から特に日本語を対象とする際，明らかな文末表現(「です」「ます」など)や特有の構文が存在するため，言語的な特徴を踏まえ文境界を推定する研究が報告されている．例として，話し言葉の係り受け解析と文境界推定の相互作用による高精度化[2]，対話音声を対象とした統計的言語モデルによる手法[3]などが挙げられる．さらに，F0 など韻律や非言語情報を用いた推定[4]，節境界検出プログラム CBAP を利用した節境界判定[5]といった研究も報告されている．

さらに，近年特にサポートベクターマシン(SVM)を用いた文境界推定が多く見受けられる．隣接文節間の係り受け情報に着目したチャンキング[6]では，SVM に与える素性を前後 3 形態素の単語情報(表層表現，読み，品詞情報)や文節内の主辞・語形の単語情報を用い，係り受け情報などを素性として用いている．また，韻律の素性を用いた話し言葉の節・文境界推定[7]では，前後 3 形態素の単語情報(表層表現，読み，品詞情報)などに加え，話している際のトーンといった韻律情報を素性として加えている．しかし，このように形態素の単語情報を使用する場合，素性として用いる発話末の単語情報の認識を誤ってしまうと正しく文境界を検出することができない．

そこで本研究では，文単位で分割したテキストで学習した言語モデルを用いて，発話末直前の形態素の品詞情報に加え，各形態素の単語信頼度を素性とした文境界の検出手法を提案する．本研究では『日本語話し言葉コーパス(CSJ)』の講演音声を対象に，文単位で分割したテキストで言語モデルを学習することで発話末に出現しやすい単語を認識しやすくし，その際の単語信頼度を導入することで文境界の検出精度の向上を目指す．

† 同志社大学
Doshisha University

2．CSJ における節境界

節とは，述語を中心としたまとまりであり，統語的にも意味的にもある程度完結した単位である．CSJ では，この節の終端境界を節境界として節境界直後の切れ目の大きさという観点から「絶対境界」，「強境界」，「弱境界」の 3 つのレベルに区分し，それぞれの節境界ラベルを与えている[8]．絶対境界は，形式上明示的な文末表現で，「言いました」，「以上です」などが相当する．強境界は，発話の大きな切れ目として考えられる従属節で，「けれども」，「が」などが相当する．弱境界は，通常は発話の切れ目になることはないと考えられる従属節で「とか」，「ので」などが相当する．これらの節境界に分類されるものを表 1 に示す．本稿では絶対境界を文境界とみなして実験を行った．

表 1: CSJ における節境界ラベル

［絶対境界］	＊デフォルト境界 文末，文末候補，と文末
/強境界/	＊デフォルト境界 並列節ガ，並列節ケド，並列節ケドモ，並列節ケレド，並列節ケレドモ，並列節シ
〈弱境界〉	タリ節，タリ節−助詞，テカラ節，テカラ節−助詞，テハ節，テモ節，テ節，テ節−助詞，トイウ節，トカ節，トカ節−助詞，ノニ節，ヨウニ節，フィラー文，引用節，引用節−助詞，引用節トノ，感動詞，間接疑問節，間接疑問節−助詞，条件節タラ，条件節タラバ，条件節ト，条件節ナラ，条件節ナラバ，条件節レバ，並列節ダノ，並列節デ，並列節ナリ，理由節カラ，理由節カラ−助詞，理由節カラニハ，理由節ノデ，連用節，連体節テノ

3．文境界モデルから得られた単語信頼度を考慮した文境界検出

SVM に与える文境界検出のパラメータとして，本研究では文境界単位で分割したテキストで学習した言語モデルを用いて認識を行い，得られた発話末からの 2 形態素の品詞情報と単語信頼度を用いた．なお，ここで述べる単語信頼度とは，Julius を使用した際，音声認識結果とともに出力される単語事後確率を指す．

3.1 従来の言語モデル

CSJ を用いて学習した講演音声認識のための標準的な言語モデル(以下，ポーズモデル)の学習データは 1000ms 以上のポーズで分割されたテキストを 1 つの発話とみなして

いる．この従来のポーズモデルにおけるテキスト分割法の例を図 1 に示す．本研究では，このポーズモデルを従来法とする．

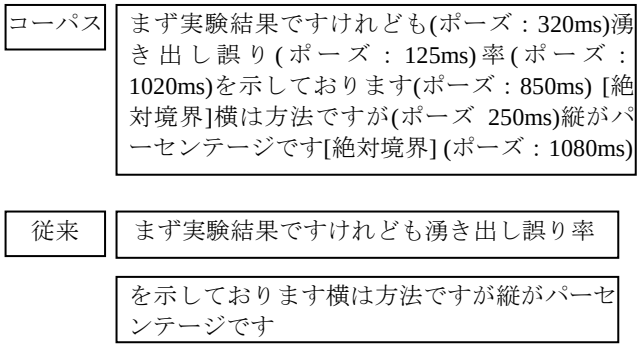

図 1: 従来のテキスト分割方法

3.2 文境界単位で学習した言語モデル

提案手法では絶対境界で分割した学習データについて考える．絶対境界で分割することで確実に発話末が文末表現になり，これにより作成した言語モデル(以下，文境界モデル)を用いて音声認識を行うと言語的な制約から文末表現が認識されやすくなる．そのため，発話末が文末表現であった場合，認識精度の向上が見込まれ，それに伴って文境界検出の向上が期待される．この提案の文境界モデルにおけるテキスト分割法の例を図 2 に示す．

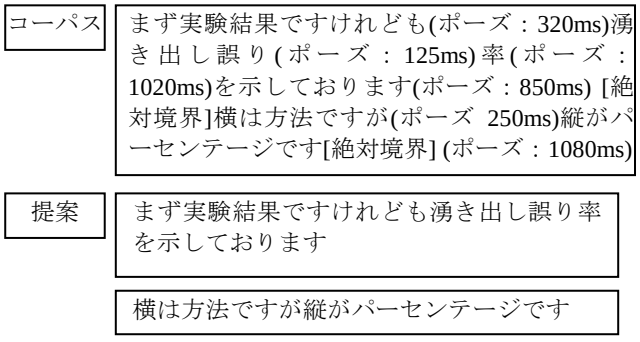

図 2: 提案するテキスト分割方法

3.3. 品詞情報と単語信頼度を利用した文境界検出

提案手法では文境界単位で学習した言語モデルにより音声認識を行い，発話末直前の 2 形態素の品詞情報に加えて，各形態素の単語信頼度を SVM の素性にした．つまり，発話末の直前 2 形態素の品詞情報と単語信頼度の計 4 個のパラメータを素性として用いた．

まず，学習データの発話を 1000ms 以上のポーズで分割して音声認識を行う．その結果から，発話末直前の 2 形態素の品詞情報と単語信頼度を抽出し，SVM により文境界のモデルを学習する．品詞情報は，形態素解析ソフト Mecab[9]を利用し認識結果文から得た．また，SVM に与える素性を特徴空間上で扱うには与える素性を数値化する必要があるため，品詞情報を数値に変換する作業を行った．例えば，格助詞は 1，感動詞は 24，名詞は 28 といったように品詞情報は全 33 種類でそれぞれに任意の数字を割り振った．なお，この中には動詞であれば動詞/連用形といったように活用形に分けた品詞も含んでいる．文境界に出現しやすい品詞情報は助動詞/終止形，終助詞，動詞/終止形であった．評価データにおいては，200ms のポーズで分割した発話を音声認識し，発話末直前の 2 形態素の品詞情報と単語信頼度を素性として，SVM により文境界かどうか判別する．

文境界モデルでは，文境界単位で分割したテキストで言語モデルを学習しているため，文末での認識精度が高くなることから単語信頼度は高くなると考えられる．また，言語モデルを学習する際に文末は必ず文境界に対応しているため，文末でない場合は認識精度が低くなることから単語信頼度は低くなると考えられる．本研究で用いたデータの発話末における形態素の品詞情報と単語信頼度の例を図 3 に示す．

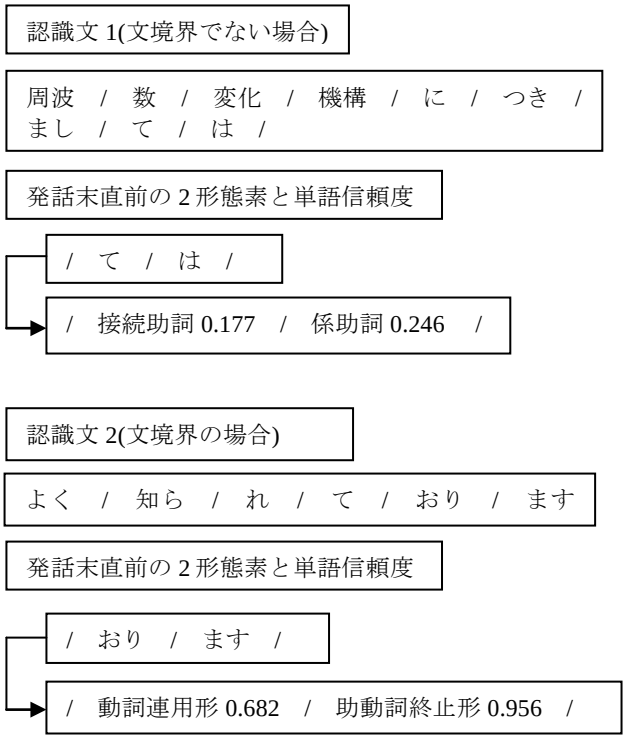

図 3: 発話末直前の形態素の品詞情報と単語信頼度の例

5. 評価実験

5.1 実験条件

本実験で使用した音響モデルは，CSJ 付属の音響モデルで，これは混合連続正規分布 HMM(対角共分散)であり，HTK で作成されている．音素毎に 3 状態 left-to-right HMM(飛び越し遷移なし)で，音素環境依存(状態共有トライフォン)でモデル化している．その際，決定木に基づく状態共有を行い，状態 3000 のモデル(16 混合)を学習している．学習データは，学会講演のなかで，男性話者 787 講演，情勢話者 166 講演の計 953 講演である．

言語モデルの学習データは，評価データに用いた学会講演 10 講演を除く，CSJ164 講演(学会講演 57 講演，模擬講演 107 講演)を用いた．言語モデルの作成には Palmkit を利用した．言語モデルを作成する際に，3 回以上出現した

形態素で構成(カットオフは 2)し，2 回以下の場合は <UNK>(未知語)として扱っている．また，言語辞書の発音エントリ総数(言語辞書内の単語数)は 6900 である．

従来の言語モデルであるポーズモデルは CSJ 付属の言語モデル作成法にのっとり，学習データの分割は 1000ms 以上のポーズで行い言語モデルを作成した．一方，提案手法である文境界モデルは学習データの分割を CSJ の節境界のうち絶対境界のラベルが付与されている部分で分割した言語モデルである．つまりどんなに長いポーズが挿入されていても絶対境界となる単語が出現していなければ発話を区切らない．

デコーダには Julius(Ver4.1.4)[10]を利用している．SVM の学習に用いたデータは発話末直前 2 形態素を対象としている．従来法，提案法をそれぞれポーズモデル，文境界モデルを用いて 1000ms のポーズで区切った 164 講演の音声を認識し算出された単語信頼度を用いた．品詞情報は認識結果に対して形態素解析を行い，任意の数値に割り振った．

評価データには CSJ に収録されている 10 講演を用いた．1 講演あたり 7 分という短いものから 26 分という長いものまで幅広く選んだ．データは 16bit で量子化，16kHz でサンプリングされている．特徴量は各フレーム毎に MFCC(12 次元)，⊿MFCC(12 次元)，⊿Power(1 次元)を計算し，計 25 次元の特徴ベクトルを求めている．なお，入力音声は予め 200ms 以上のポーズで分割しているものを利用する．また，個人名など音声にノイズが重畳している部分については評価対象から除外している．

ポーズモデルと文境界モデル作成時の文数と文境界数を表 2，本実験で用いる学習データの発話の数を表 3 に示す．

表 2: ポーズモデルと文境界モデルの発話数と文境界数

言語モデル	文数	境界数
ポーズモデル	7019	3124
文境界モデル	9729	9729

表 3: 学習データの発話数

データ総数	発話末が境界	発話末が境界でない
7019	3124	3895

表 2 から，ポーズモデルでは 1000ms のポーズごとにテキストを分割しているため，文数と境界数は異なっている．それに対して，文境界モデルでは絶対境界でテキストを分割しているため，文数と文境界が一致している．

5.2 音声認識結果

本実験で作成したポーズモデルと文境界モデルを利用し，学習データと評価データの認識を行った.この結果を表 4 に示す.

表 4: ポーズモデルと文境界モデルの認識精度

言語モデル	学習データ	評価データ
ポーズモデル	66.6%	63.1%
文境界モデル	66.7%	62.4%

表 4 より文境界モデルはポーズモデルと同等の認識精度であることがわかる．また，評価データに対して，文境界とそうでないときの発話末の単語認識精度を求めた結果を表 5 に示す.

表 5: ポーズモデルと文境界モデルの発話末単語認識精度

言語モデル	発話末が境界	発話末が境界でない
ポーズモデル	64.2%	68.8%
文境界モデル	65.7%	64.0%

表 5 よりポーズモデルに比べ，文境界モデルは文末が境界と一致する際，認識精度が 1.5%上昇していることがわかる．すなわち，文境界モデルを用いることで発話末が文末表現になりやすいという特徴から発話末の単語信頼度に影響があると考えられる．なぜなら，元々発話末が文末表現の箇所に対しては文境界モデルの特徴に合致しているため，認識の際の展開候補単語数は少なくなり本研究で利用する単語信頼度は高い数値が得られると考えられるからである.

5.3 文境界の検出結果

対象音声の発話末が境界で区切れている場合と境界で区切れていない場合に対して文境界検出を行った．発話末が境界である場合の文境界の判別精度を表 6，発話末が境界でない場合の文境界の判別精度を表 7 に示す．なお，対象が書き起こしの素性は発話末とその直前の 2 形態素の品詞情報を使用している．また，素性欄の単語信頼度(1)は，発話末の形態素のみに対する単語信頼度を用いた場合の結果である．それに対して単語信頼度(2)は，発話末直前 2 形態素の単語信頼度を用いた場合の結果である.

表 6: 発話末が文境界のときの境界判別精度

対象	素性	正答率	適合率	F 値
書き起こし	品詞情報のみ	95.3% (716/751)	72.9 (716/982)	82.6
ポーズモデル	品詞情報のみ	81.6% (613/751)	47.0% (613/1303)	59.7
	品詞情報+単語信頼度(1)	75.9% (570/751)	70.9% (570/804)	73.3
	品詞情報+単語信頼度(2)	76.2% (572/751)	70.0% (572/817)	73.0
文境界モデル	品詞情報のみ	78.6% (590/751)	58.8% (590/1004)	67.2
	品詞情報+単語信頼度(1)	70.7% (531/751)	76.2% (531/697)	73.3
	品詞情報+単語信頼度(2)	68.0% (511/751)	81.8% (511/625)	74.3

表 7: 発話末が文境界でないときの境界判別精度

対象	素性	正答率	適合率	F 値
書き起こし	品詞情報のみ	88.2 (1979/2245)	98.3 (1979/2014)	92.9
ポーズモデル	品詞情報のみ	69.3% (1555/2245)	91.8% (1555/1693)	79.0
	品詞情報＋単語信頼度(1)	89.6 (2011/2245)	91.7 (2011/2192)	90.6
	品詞情報＋単語信頼度(2)	89.1% (2000/2245)	91.8% (2000/2179)	90.4
文境界モデル	品詞情報のみ	81.6% (1831/2245)	91.9% (1831/1992)	86.4
	品詞情報＋単語信頼度(1)	92.6 (2079/2245)	90.4 (2079/2299)	91.5
	品詞情報＋単語信頼度(2)	94.9% (2131/2245)	89.9% (2131/2371)	92.3

表 6，表 7 の結果からまず，SVM に与える素性を品詞情報のみとした際，ポーズモデルより文境界モデルを用いることで F 値の向上が見られ，提案手法が従来手法よりも文境界かどうかの検出で書き起こし結果により近い精度が出ている．特に，表 7 の文境界でない判別では提案手法は書き起こしの結果とほぼ同等の結果が得られている．この結果から文の境界検出に適する言語モデルは文境界モデルの方であると考えられる．

また，いずれの表からも SVM に与える素性に品詞情報のみを与える時と，加えて単語信頼度を利用した時を比較すると単語信頼度を利用した場合の方が F 値の向上が見られた．これは発話末が境界である時と境界でない時の認識精度の差が影響していると考えられる．表 5 から見ても発話末が境界のときと境界でない時の発話末の単語に対する認識精度に差がある．つまり，文境界であれば展開候補単語数が少なくなり単語信頼度は高くなる．そして，文境界でなければ展開候補単語が多くなり単語信頼度は低くなるという特徴が境界判別に適合したと考えられる．特に,文境界モデルを用いた際，元々発話が文末表現である箇所で区切られていた場合においては言語モデルの特徴から認識がしやすくなるため品詞情報のみの場合でもポーズモデルと比べると F 値の向上が見られた．この結果から単語信頼度の素性としての有効性が明らかになったと考えられる．

さらに，文境界モデルの利用においては，単語信頼度を発話末の形態素のみに利用するよりも，発話末 2 形態素を利用した場合がより F 値の向上が見られた．

以上の結果から，従来のポーズモデルに比べ文境界モデルを利用すること，さらに発話末の品詞情報に加えて単語信頼度を用いるという提案手法は文境界の検出に有効であることがわかった．

6. おわりに

本研究では，日本語話し言葉コーパス(CSJ)の講演音声を対象に音声認識結果に対する文境界の検出精度向上を目的として，文境界単位で学習した言語モデルを用いて得られた発話末の形態素の単語信頼度を SVM の素性とする手法を提案した．この提案手法を用いることで，従来法をポーズで区切った文から学習した言語モデルを用い SVM に与える素性を品詞情報のみとした際，文境界の検出精度は F 値 59.7 であるのに対して，提案手法では F 値 74.3 となり提案手法の有効性が明らかになった．

今後の課題として，表層情報，読み，品詞情報などを用いた従来法との比較実験，提案手法において文末の認識精度がどれほど文境界検出の性能に影響するかを分析する必要がある．また，ポーズを伴わない文境界の検出法や，複数の認識器の併用による文境界の検出手法についての検討を行う予定である．

謝辞

本研究は，科研費基盤研究(B)(21300066)の助成を受けたものである．

参考文献

[1] 野村和弘，河原達也，堂本修司，“講義の自動アーカイブ化のための韻律情報を用いた講義音声の文境界の抽出”，電子情報通信学会信学技報，SP98-80，pp.17-24 (1998).

[2] 下岡和也，内元清貴，河原達也，井佐原均，“話し言葉の係り受け解析と文境界推定の相互作用による高精度化”，自然言語処理学会，NLP-12(3) pp.3-17 (2005).

[3] 中嶋秀治，山本博史“音声認識過程での発話分割のための統計的言語モデル”，情報処理学会論文誌，Vol.42(11) , pp.2681-2688 (2001).

[4] 小橋修一，山下洋一，“音声要約のための韻律情報を用いた文境界の自動決定”，日本音響学会秋季研究発表会講演論文集，3-7-8 (2005) .

[5] 丸山岳彦，柏岡秀紀，熊野正，田中英輝，“日本語節境界検出プログラム CBAP の開発と評価”，自然言語処理学会，NLP-11(3), pp.39-68 (2004).

[6] 西光雅弘，河原達也，高梨克也，“隣接文節間の係り受け情報に着目した話し言葉のチャンキングの評価”，情報処理学会研究報告，SLP-61(4), pp.19-24 (2006).

[7] 尾嶋憲治，秋田祐哉，河原達也“局所的な係り受けと韻律の素性を用いた話し言葉の節・文境界推定”，情報処理学会研究報告，SLP-67(3), pp13-18 (2007).

[8] 高梨克也，内元清貴，丸山岳彦，“『日本語話し言葉コーパス』における節単位認定”，http://www.kokken.go.jp/katsudo/seika/corpus/public/manuals/asr.pdf

[9] http://mecab.sourceforge.net/

[10] 李晃伸,“大語彙連続音声認識エンジン Julius”，電子情報通信学会情報・システムソサイエティ誌, Vol. 13, No. 4, (2009-2).