

Internal Use Only (非公開)

TR-SLT-0005

信頼度尺度による音声認識誤りの検出と
用例データベースに基づく音声認識誤りの訂正
に関する検討

Mis-Recognition Detection Using Confidence Measures,
and Mis-Recognition Correction Based on Expression Corpus

沖本 純幸	山本 博史
Yoshiyuki Okimoto	Hirofumi Yamamoto
匂坂 芳典	菊井 玄一郎
Yoshinori Sagisaka,	Genichiro Kikui

2002年2月28日

概要

本レポートでは、音声認識に不可避な認識誤りの問題に対する取り組みとして、認識誤り個所の検出と、これの訂正に関する検討について報告する。認識誤りの検出問題については、これまで種々の尺度およびその併用方法が提案されているが、有効な尺度、有効な併用方法についての検討結果を報告する。また、認識誤りの訂正問題については、事前に構築した表現用例データベースを知識源として認識誤り個所を訂正する方法について、検討、提案する。また最後に、これらの実験のために構築した各種ツール類の利用方法を補足資料として掲載する。

(株) 国際電気通信基礎技術研究所
音声言語コミュニケーション研究所
〒619-0288 「けいはんな学研都市」 光台二丁目2番地2 TEL: 0774-95-1301

Advanced Telecommunication Research Institute International
Spoken Language Translation Research Laboratories
2-2-2 Hikaridai "Keihanna Science City" 619-0288, Japan
Telephone: +81-774-95-1301
Fax : +81-774-95-1308

©2002 (株) 国際電気通信基礎技術研究所
©2002 Advanced Telecommunication Research Institute International

目次

1	信頼度尺度による音声認識誤りの検出	1
1.1	はじめに	1
1.2	信頼度尺度	1
1.2.1	音響的要因による信頼度尺度	2
1.2.2	言語的要因による信頼度尺度	2
1.2.3	探索,あるいは音響・言語両要因による信頼度尺度	4
1.3	種々の信頼度尺度の比較	5
1.3.1	音声認識システム	5
1.3.2	評価尺度	5
1.3.3	実験条件	6
1.3.4	実験結果	6
1.4	認識誤り検出の傾向比較	7
1.4.1	誤り原因の同定	7
1.4.2	実験条件	7
1.4.3	実験結果	7
1.5	複数尺度の組合せによる認識誤り検出	9
1.5.1	信頼度尺度の相関係数	9
1.5.2	3尺度の組合せによる比較実験	10
1.5.3	多尺度の組合せによる評価	11
1.6	まとめ	12
	参考文献	12
2	旅行会話基本表現コーパスを用いた認識誤り訂正の検討	14
2.1	はじめに	14
2.2	提案方式	14
2.2.1	本研究における認識誤り箇所	15
2.2.2	代替候補の生成	15
2.2.3	最適候補の選択	15
2.3	提案方式の評価	16
2.3.1	実験条件	16
2.3.2	実験結果	16
2.4	音素距離の検討	17
2.5	文類似度の評価	18
2.5.1	文類似度	18
2.5.2	実験結果	18
2.6	誤り訂正結果の分析	19
2.6.1	誤り長ごとの評価	19
2.6.2	品詞ごとの評価	19
2.7	まとめ	20
	参考文献	20

A	音声認識誤りの検出・訂正ツールについて	22
A.1	はじめに	22
A.2	音声認識誤り検出のフロー	23
A.2.1	信頼度尺度計算のための前準備	23
A.2.2	信頼度尺度計算	24
A.2.3	複数の信頼度尺度のマージ	25
A.2.4	評価対象外単語の特定	25
A.2.5	学習機械用データの生成	25
A.2.6	学習機械のトレーニング	25
A.2.7	認識誤り検出	26
A.2.8	生成ファイルのフォーマット	26
A.3	音声認識誤り訂正フロー	29
A.3.1	用例データベースの構築	29
A.3.2	類似用例文検索	29
A.3.3	認識誤り訂正候補の生成	30
A.3.4	誤り訂正の集計	30
A.3.5	生成ファイルのフォーマット	30
A.4	各ツールについて	32
	CalcCM.pl	33
	CMMerge.pl	35
	ReFlag.pl	36
	ExtMeasure.pl	37
	ErrorDetect.pl	39
	LinearDisc.pl	41
	LinearDisc2.pl	42
	MkDecisionRule.pl	43
	MkMLPProj.pl	44
	MkExampleDB.pl	45
	ChkExample.pl	47
	MkHypothesis.pl	50
	CorrectionEval.pl	51
A.5	信頼度尺度計算モジュールについて	53
A.5.1	CMNull	53
A.5.2	CMname	53
A.5.3	CMPOS	53
A.5.4	CMTagging	53
A.5.5	CMClassID	53
A.5.6	CMWordLeng	53
A.5.7	CMScore	53
A.5.8	CMBackScore	54
A.5.9	CMBasic	54
A.5.10	CMDivideMerge	54
A.5.11	CMEndState	55
A.5.12	CMStripHypo	55
A.5.13	CMStripHypo3	55
A.5.14	CM_Hypo_Dist	55
A.5.15	CMPosteriori	56
A.5.16	CMPosteriori2	56
A.5.17	CMClassHypo	56
A.5.18	CMCompArc	57
A.5.19	CMPhDuration	57

A.5.20 CM2gramCount	57
A.5.21 CMLangEntropy	57

第 1 章

信頼度尺度による音声認識誤りの検出

1.1 はじめに

近年多くの大語彙連続音声認識システムが提案され、そのいくつかのものは、商用アプリケーションとして市場に出回るようになってきた。しかしこれらのシステムも、完全にエラーなしとなった訳ではない。このため、音声認識技術の実用的な応用を考えると、認識誤りに対する対策は必要不可欠である。

一般に、音声認識誤りに対する対策としては、2つのアプローチが考えられる。1つは音声認識システムそのものの性能を上げるためのアプローチであり、これには音響・言語モデルの改良、あるいは探索手法に対する改良などが挙げられる。もう1つのアプローチは、音声認識器からの出力を検証することによって、認識誤りの低減を目指すものである。このようなアプローチが有効であると考えるのは、局所的な認識誤りに対して、その前後の広範囲な言語的制約など探索の過程で利用してない情報を用いることで、認識誤りを訂正ができると考えられるためである。後者のアプローチを採る場合、音声認識器出力に含まれる認識誤りを検出する技術が重要となる。

我々は、現在音声翻訳器の開発を行っており、その入力フロントエンドとしての音声認識の性能を上げるため、上に述べたように後处理的に音声認識誤り箇所を修正する方法を検討している。このような目的にしたがって、本論文では音声認識誤りの検出法に関する検討について述べる。

音声認識誤りの検出は、一般的には音声認識の計算過程で得られる種々の尺度を用いて行なわれる。音声認識結果の確からしさを表わすこれらの尺度は、信頼度尺度 (confidence measure) と呼ばれて近年多くの提案・検討がなされている。これらの研究の目的は、音声認識誤りの検出 [8, 1] に限らず、いくつかの目的の下で検討がなされている。たとえば、辞書外語彙の発話の検出を目的としたもの [7] や、信頼度情報を考慮した探索による性能の改善 [15, 3]、対話システムへの応用 [4]、などが挙げられる。こうした研究の中では様々な信頼度尺度が提案されているが、それは大別して音響的な意味を持つ尺度、言語的な意味を持つ尺度、探索 (あるいは音響・言語両方) に関する意味を持つ尺度、に分類することができる。音響的な尺度としては、例えば対立候補との音響的尤度比を利用するもの [7] などがあり、言語的な尺度としては、言語スコアそのものや back-off の発生回数を利用したもの [1]、探索に関する尺度としては、単語グラフから得られる事後確率を用いるもの [2] などが挙げられる。こうした尺度を用いた先行研究の中で多くの成功した事例では、複数の尺度を学習機械でトレーニングし、判別するという方法が採られている。これは、一言に音声認識誤りと言ってもその原因は非常に多様であり、それらについての十分なモデル化が困難であるためと考えられる。またこのために、上述したような先行研究においても、これら信頼度尺度の効果についての検討は不十分である。

そこで我々は、実験を通していくつかの信頼度尺度を用いた認識誤り検出について、性能、傾向についての分析を行ない、認識誤りの検出に有効な信頼度尺度を明らかにする。

1.2 信頼度尺度

本研究が目的とする音声認識誤りの検出は、音声認識器から出力される各単語について、当該単語が正解 (1) か認識誤り (0) かを判別する、2値判別の問題と考えることができる。この判別の根拠となる尺度が、信頼度尺度である。本論文では、先行研究も踏まえて本論文で検討する種々の信頼度尺度について述べる。先にも述べたように、音声認識の誤り検出に利用されるこれら信頼度尺度は、音響的な要因によると考えられる尺度、言語的な要因によると考えられる尺度、探索、あるいは音響と言語の両要

困によると考えられる尺度に大別できる。以下では信頼度尺度について、上記3つのカテゴリに分けて述べる。

1.2.1 音響的要因による信頼度尺度

音響尤度 (AL)

音響スコアそのものは、ある程度単語の正解率と相関を持つ。とりわけスコアが低い時には、正解率も下がるものと考えられる。そこで、音響スコアを、単語のフレーム長で正規化した値を利用する。

音響尤度比 (ALR)

多くの辞書外発話の検出に関する研究で提案されているように [7]、音響スコアをそのまま使うより、正規化あるいは事後確率化の方が検出精度の向上が見込まれる。ここでは、1位候補単語の音響スコアのフレーム平均と、同じ時間窓周辺に表われる2位候補単語の、音響スコアのフレーム平均の差を利用する。

正規化音響尤度 (NAL)

ALR同様に、単語の音響スコアの正規化を行なう。ここでは2位候補の音響スコアを利用するのではなく、並列に走らせたモーラタイプライタを利用して、1位候補単語と同じ時間窓でのモーラタイプライタの音響スコアの差を、フレームで平均した値を利用する。

音素仮説数 (PH)

後述する競合単語仮説数 (WH) のように、尤度幅一定のビーム探索において、ビーム内に残る仮説の数は1位候補の正解率とある程度の相関があると考えられる。ここでは、単語の時間窓内で出現するモーラの種類の数を用いる。これは、モーラタイプライタの計算過程で出現するモーラ候補をグラフ化した“モーラグラフ”を利用して、単語時間窓内のモーラ種類数をカウントし、フレーム数で平均したものを利用する。

音素距離 (PD)

音声認識過程では、観測された音響系列に最も適合する単語の音響モデルを、言語モデルの制約の中から選択する。このように考えた場合、語彙の制約を持たないモーラタイプライタによる音韻系列と、1位候補単語の音韻系列の類似性は、音韻的なモデルの整合性を示す尺度となっていると見ることができる。そこで、文献 [?] で述べられている尺度同様に、1位候補単語の音素列とモーラタイプライタの音素列の編集距離を、単語の音素数で平均した値を利用する。

音素持続時間確率 (PHD)

一般に音韻の持続時間は、一定の分布を持つと考えることができる。したがって、この分布傾向から極端に逸脱した持続時間を持つ音韻は、モデルがうまく整合していないことを表わしていると予想することができる。このような意味を持つ尺度として、事前に求めた各音素の持続時間確率分布を利用し、候補単語を構成する各音素の持続時間確率について、その最小確率の対数値を利用する。

単語音素長 (WDL)

1モーラ単語 (助詞など) のような短単語では、音響的な差がつきにくく誤認識することが多いと考えられる。そこで単語の長さを表わす尺度として、単語に含まれる音素の数を利用する。

1.2.2 言語的要因による信頼度尺度

言語スコア (LP)

AL同様に、言語スコアそのものによって、言語的な接続の不自然さを検出できると考えられるため、これを利用する。

言語スコア差 (LPR)

ALRと同様であり、1位候補の言語スコアと2位候補の言語スコアの差によって、誤りを検出しようとするものである。

クラス遷移スコア (CP)

我々は言語モデルとして、クラス Ngram 言語モデル [12] を用いている。これは、単語履歴の条件付き確率を次のように近似する。

$$\begin{aligned} P(w|w_{i-1}, w_{i-2}, \dots, w_{i-n}) \\ \approx P(w|C) \times P(C|C_{i-1}, C_{i-2}, \dots, C_{i-n}) \end{aligned} \quad (1.1)$$

ここで、 $P(C|C_{i-1}, C_{i-2}, \dots, C_{i-n})$ は、履歴中の各単語の属するクラスの遷移確率を表わしており、言語の構文的な制約を示すと考えられる。そこで、構文的な不自然さを表わす尺度として、クラス遷移確率の対数値を利用する。

単語出力スコア (WE)

クラス Ngram 言語モデル式 (1.1) において、クラスから単語が出力される確率 $P(w|C)$ は、単語選択の妥当性を示す尺度となると考えられる。よって、単語出力確率の対数値を利用する。

後向き言語スコア (BLP)

後向きの言語モデル、すなわち後続単語系列から先行する単語の確率を予測するモデルを用いて、言語の妥当性を示す尺度とすることができる。ここでは、後向き言語モデルによる言語スコアをそのまま利用する。

後向きクラス遷移スコア (BCP)

CPと同様に、後向き言語モデル (クラス Ngram) によるクラス遷移スコアを、構文的な妥当性を示す尺度として利用する。

後向き単語出力スコア (BWE)

BE同様に、後向き言語モデル (クラス Ngram) における単語出力確率も利用することができる。通常のクラス Ngram モデルにおいては、単語出力確率は前向き、後向きで同じ値となるが、我々の用いる言語モデルでは接続の方向性を考慮したモデル化をおこなっているため [12]、前向き、後向きで異なる値となる。従って、後向き単語出力スコアも利用する。

単語遷移パープレキシティ (TP)

Ngram 言語モデルにおいては、先行する単語から全ての後続単語への遷移確率によるエントロピーが大きい場合、後続単語の予測が困難となり、認識誤りが発生しやすくなると考えられる。そこで全語彙 V に対して、先行する単語 w_{i-1} のパープレキシティを以下のように計算し、これを尺度として用いる。

$$\begin{aligned} CM &= 2^{H(w_{i-1})} \\ H(w_{i-1}) &= - \sum_{w_i \in V} P(w_i|w_{i-1}) \log P(w_i|w_{i-1}) \end{aligned} \quad (1.2)$$

ただし、語彙 V の全ての単語に対してこれを計算するのは現実的ではないので、探索時にビーム内に残った単語だけを用いて、(それ以外の単語は、確率が十分低いと仮定して) 計算を行なう。

トレーニングサンプル数 (TS)

Ngram 言語モデルにおいて、学習時に出現しなかった単語並びに対しては、通常スムージングによって微小な確率が与えられる。しかし、このような値は理論的根拠に乏しく、誤認識につながるケースも見受けられる。そこで、学習時のクラス Ngram の出現回数によって、言語確率の統計的信頼度を表わす尺度として利用する。

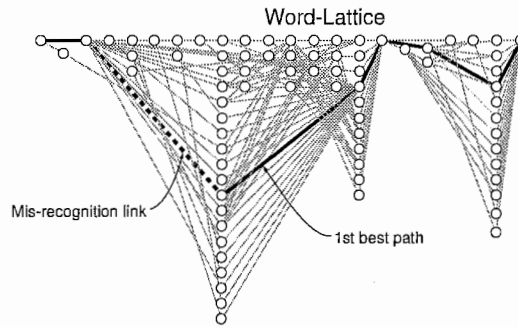


図 1.1: 認識誤りを含む単語グラフの例

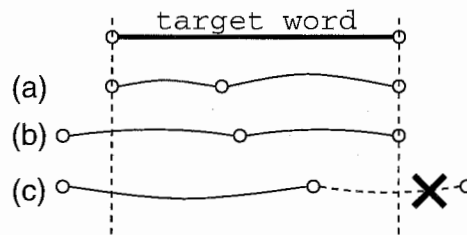


図 1.2: 単語競合条件: (a) 注目単語の時間区間に重なる単語または単語系列, (b) 時間窓からはみ出しは認める, (c) 音韻系列で比較した場合に余剰と見なされた単語は削除する

1.2.3 探索,あるいは音響・言語両要因による信頼度尺度

トータルスコア差 (TPR)

一般的に, ある候補単語が同一時間窓内で競合する他の単語より非常に大きなスコア差を持っている時, この単語は正解していることが多い。逆に同程度のスコアの候補が多く現われる場合には, 誤認識しやすい。これらは, 観測された現象に対してモデルの識別能力が十分であれば, 異なる現象が大きなスコア差となって識別されるが, 不十分な場合には類似したスコアになってしまうためと考えることができる。そこでこのような状況を捉えるため, 1位の候補と下位候補とのスコアの差を用いる。ここでは, 同一の時間窓内に存在する1位候補と2位候補のスコア(音響スコア+言語スコア)の差を用いる。

単語始端分岐数/終端合流数 (NSD/NEM)

上述したようなモデルの識別能力が十分であるか否かという問題は, 尤度幅一定のビーム探索においてビーム内に残る仮説の数によっても推測することができる。すなわち, スコア差が開いており, 正解していると推測される場合には, ビーム内の仮説は少なく, 逆に類似スコアの候補が多く表われて, 認識誤りの発生が多いケースでは, ビーム内の仮説は多くなると考えられる。図 1.1には, 音声認識システムの実出力候補の中間表現である単語グラフを用いて, 認識誤りが発生した場合の競合候補の数の増加の様子を一例として示している。

このようなビーム内の仮説の数を捉える尺度として, いくつかの尺度が提案されている [8]。ここでは, 注目単語の始端時刻に開始する単語仮説の数, および終端時刻に終端する単語仮説の数を利用する。

競合単語仮説数 (WH)

上述の NSD や NEM では, 時間的に互いにオーバーラップする単語仮説の数を直接的にカウントしている訳ではない。また, 同一の単語が時間的に少しづつずれながら何度も出現するような状況

では、うまく仮説数をカウントすることができない。そこで、直接的にオーバーラップする単語の種類をカウントする方法を提案する。

我々の方法では、注目単語の時間区間を通過する単語グラフ上の全ての部分パスを、単語ID系列でマージした数を利用するものとした。この時、音響モデルのあいまい性によって候補の出現が時間的にゆらぎ、始端、終端位置に隣りの単語が入り込むことがある。このため、音素系列の比較によってこれら余分なリンクを削る。これを図1.2に図示する。

単語事後確率 (WPP)

単語事後確率は、近年、信頼度尺度として注目され、盛んに研究されている [2]。単語事後確率は観測された音響系列 X に対して、時刻 $t_s \sim t_e$ の時間窓の単語が w である確率 $p(w, t_s, t_e | X)$ として以下のように求める。

$$\begin{aligned} p(w, t_s, t_e | X) &= \sum_{W_a} \sum_{W_e} p(W_a, w, W_e | X) \\ &= \frac{\sum_{W_a} \sum_{W_e} p(X | W_a, w, W_e) p(W_a, w, W_e)}{p(X)} \end{aligned} \quad (1.3)$$

候補表現として単語グラフを用いている場合、上式の分子は、単語グラフの始端ノードから終端ノードへ向かうの全てのパスのうち、注目単語に対応するリンクを通過するパスの文確率を合計したもので近似できる。また分母については、単語グラフ中の全てのパスの文確率を足し合わせたもので近似できる。すなわち単語事後確率は、単語グラフ中の全てのパスの中で注目するリンクが選ばれる確率を計算するものである。ただし、我々の単語グラフでは同一の単語、同一の時間窓であっても異なるコンテキストであれば、異なるリンクを割り当てているため、実際には、上述の **WH** と同様の競合条件で一致する単語の事後単語確率どうしを足し合わせることによって、目的単語の事後単語確率として計算する。

1.3 種々の信頼度尺度の比較

前節に述べた 21 種類の信頼度尺度による、認識誤りの検出能力を比較する実験を行なう。実験では、各尺度をそれぞれ単独で用いて、認識結果の各単語ごとにその正解・誤りを推定する。

1.3.1 音声認識システム

実験に用いた音声認識システムについて概要を述べる。本音声認識システムでは、探索は 2 パスで行なわれる。1 パス目は音素環境異存の音響モデルと、同音語をマージした言語モデルによって探索が行なわれ、2 パス目でより詳細な言語モデルによってリスクアリングされる。探索はフレーム同期で行なわれ、尤度幅一定のビームによって枝刈りが行なわれる。ビームに残った仮説は、図 1.1 に示すような単語グラフの形態で出力される。この単語グラフでは、ノードとノードを繋ぐリンクが各単語仮説に対応しており、各単語の音響スコア、言語スコア、始終端時間などが合わせて保存されている。

1.3.2 評価尺度

本実験で行なう信頼度尺度による認識単語の正解・誤りの推定には、単純な閾値による判定を行なう。すなわち、各単語ごとに計算した値に対して、これと閾値との大小によって、単語の正解・誤りを推定する。この推定された単語の正解・誤りと、あらかじめ目視によって作成された正解ラベルから決定される“真の”の正解・誤りとの適合度を比較することによって、尺度の評価を行なう。

ここで、以下に示す 2 つの値、誤り適合率と誤り再現率の間には閾値の増減に対してトレードオフの関係があることが知られている。

$$\begin{aligned} \text{誤り適合率 (ER)} [\%] &= \frac{\text{誤りと推定された誤り単語の数}}{\text{誤りと推定された単語の数}} \times 100 \end{aligned}$$

信頼度尺度	ER [%]	DR [%]
音響的尺度		
音響尤度 (AL)	21.7	80.1
音響尤度比 (ALR)	39.1	84.6
正規化音響尤度 (NAL)	31.1	82.6
音素仮説数 (PH)	23.0	80.4
音素距離 (PD)	29.6	82.1
音素持続時間確率 (PHD)	19.5	79.5
単語音素長 (WDL)	8.0	77.1
言語的尺度		
言語スコア (LP)	31.2	82.6
言語スコア差 (LPR)	35.5	83.7
クラス遷移スコア (CP)	27.2	81.5
単語出力スコア (WE)	26.7	81.2
後向き言語スコア (BLP)	20.3	77.4
後向きクラス遷移スコア (BCP)	19.2	77.2
後向き単語出力スコア (BWE)	17.9	76.8
単語遷移パーレキシティ (TP)	33.8	83.2
トレーニングサンプル数 (TS)	29.1	82.0
音響+言語または探索に関連する尺度		
トータルスコア差 (TPR)	44.4	85.8
単語始端分岐数 (NSD)	51.3	87.7
単語終端分岐数 (NEM)	48.2	86.9
競合単語仮説数 (WH)	55.0	88.5
単語事後確率 (WPP)	52.9	88.0

表 1.1: 各尺度の誤り検出比較

$$\begin{aligned} & \text{誤り再現率}[\%] \\ & = \frac{\text{誤りと推定された誤り単語の数}}{\text{誤り認識した単語の数}} \times 100 \end{aligned}$$

そこで、ここでは閾値を振って上記の2つの値が同じになる閾値を探し、その条件での誤り再現率 (= 誤り適合率) を用いるものとした。また同時に、以下に示すような判別率によっても評価を行なうものとする。

$$\begin{aligned} & \text{判別率}(\text{DR})[\%] \\ & = \frac{\text{正しく推定された単語の数}}{\text{全ての判定された単語の数}} \times 100 \end{aligned}$$

1.3.3 実験条件

本実験で用いる認識器は、音響モデルとして男性および女性の不特定話者のHMnetを用いており、言語モデルには、多重クラス2-gram[12]を用いている。探索は、尤度幅一定のビームによる枝刈りによる2パス探索を行なう。評価用のデータとしては、ATR旅行会話データベース[13, 14]から、発話数930実験では、単語Accuracyで84.2%の性能を得た。

1.3.4 実験結果

1.3.2節に述べた基準で評価した、各尺度の実験結果を表1.1にまとめる。

本実験では、種々の信頼度尺度を音響的尺度、言語的尺度、音響+言語的尺度の3つに大きく分けて比較したが、それら3つの中では、音響+言語的尺度による推定が最も良いことが、本結果の比較から確認できる。たとえば、類似した尺度である音響尤度比・言語スコア差・トータルスコア差による結果を比べても、トータルスコア差によるものが最も良いことが示されている。これは本実験で用いた音声認識器が音響モデルと言語モデルの2つのモデルに基づいて候補を生成しているため、仮に一方のモデ

ルで識別困難な入力音声であっても、もう一方のモデルで十分識別可能であれば正解候補が導ける可能性があり、このため、音響または言語的な尺度だけから判断することは困難であるためと考えられる。

また音響的尺度のみについて見ると、音響的尺度の中では音響尤度比による推定が最も良く、次いで正規化音響尤度が高い。これらは共に音響尤度正規化（近似的に音響確率の事後確率化）を行なうものであり、基準尤度を2位候補の尤度とするかモーラタイプライタによる尤度とするかの違いであり、本質的には同じ狙いによる尺度であると考えられる。また、従来より未登録語発話の検出にこのような尺度が用いられてきたが、本結果は、これらの尺度の有効性を裏付けるものとなっていると考える。また本結果では、音素持続時間確率による推定精度は非常に低い。これは、今回の実験で用いた音声は旅行対話音声であるため、比較的自然的な発話で話中の発話速度の変動が大きかったことと、音素ごとの持続時間確率のモデル化において、音素の文脈を考慮せず均一にモデル化するという粗いモデル化であったために、十分な識別ができなかったととえられる。

次に言語的尺度について比較すると、言語スコア差によるもの、次いで単語遷移パープレキシティが最も良い結果となった。言語スコア差は2つの言語現象の識別の程度を表わしていると考えれば、識別が困難な場合に認識誤りを起こしやすいと解釈できる。また、単語遷移パープレキシティも同様に、ある単語の次の単語の予測において、分岐が多い時すなわち次単語の予測が困難である時、認識誤りを起こしやすいと解釈できる。したがって、観測された現象をモデルが識別できる程度を表わす尺度が有効であると考えることができる。

最後に探索に関連する尺度では、競合単語仮説数または単語事後確率が最も良い結果であった。またこれらの尺度は、21の信頼度尺度の中でも一番良い推定結果であった。特に競合単語仮説数は、従来提案されてきた単語始端位置あるいは終端位置での分岐、合流数を用いた場合よりも良好な結果となっている。これは、1.2.3節でも述べたように、単語仮説は時間的なゆらぎを持って出現するが、これによる影響をうまく取り除くことができたためであると考えられる。また、これらの尺度がトータルスコア差を用いた場合より良い結果を示したのは、スコア差を用いる場合には、長さについて正規化する必要があると考えられるが、同一時刻に開始・終了するとは限らない2つの候補のスコアの正規化が不十分であるためと考えられる。また単語事後確率は、実際の計算においてビーム内に残ったパスのみで全体の確率の総和が1.0となるよう近似計算している。このため単語事後確率は、ある時間窓を通過するパスの総数（すなわち、競合単語仮説数）と強い関連があり、結果としてこの両者が同程度の推定結果となったと考えられる。

1.4 認識誤り検出の傾向比較

1.4.1 誤り原因の同定

信頼度尺度の特徴を明らかにするため、各信頼度尺度が捉えることのできる誤りのタイプを調べる。ここでは、誤りのタイプとして認識誤り原因の同定 [16] による分類を行なう。認識誤り原因の同定は、認識誤りの原因を音響モデル原因誤り、言語モデル原因誤り、音響・言語両モデルの誤り、探索エラーの4つに分類する。これは、認識結果の1位系列中の単語の音響スコア、言語スコアおよびトータルスコアと、正解系列中の単語の音響スコア、言語スコア、トータルスコアを比較することにより行なう。このアルゴリズムを図 1.3 に示す。この方法で原因を同定した認識誤り単語に対して、各信頼度尺度はどのタイプの誤りを検出するか実験によって調べる。

1.4.2 実験条件

本実験でも、1.3節と同じ条件で各信頼度尺度ごとに認識誤りの検出を行なう。ここでも、閾値は誤り検出率と誤り再現率が等しくなる値とする。この時、上述の誤り原因ごとに認識誤り検出の再現率を調べる。

1.4.3 実験結果

まず、音響的な尺度と言語的な尺度の代表として、音響尤度比（ALR）と言語スコア差（LPR）について比較した結果を図 1.4 に示す。

この結果から、音響尤度比を用いた場合は、言語モデル原因の誤り再現率が音響モデル原因の誤り再現率より高く、逆に言語スコア差を用いた場合には、音響モデル原因の誤り再現率が言語モデル原因の

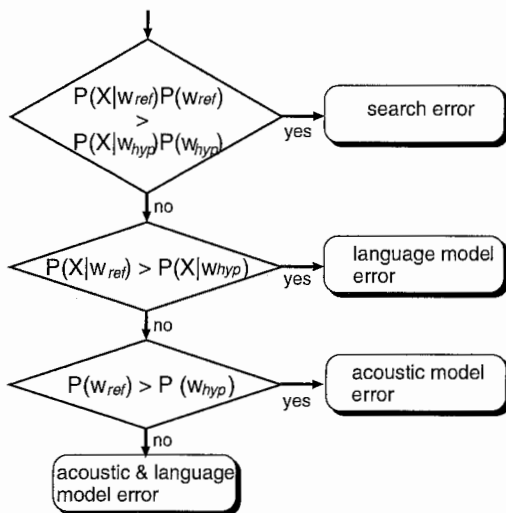


図 1.3: 誤り原因の同定: 図中 $P(X|W_{ref}), P(W_{ref})$ は正解の音響, 言語スコアをそれぞれ意味しており, $P(X|W_{hyp}), P(W_{hyp})$ は 1 位候補の音響, 言語スコアを意味している

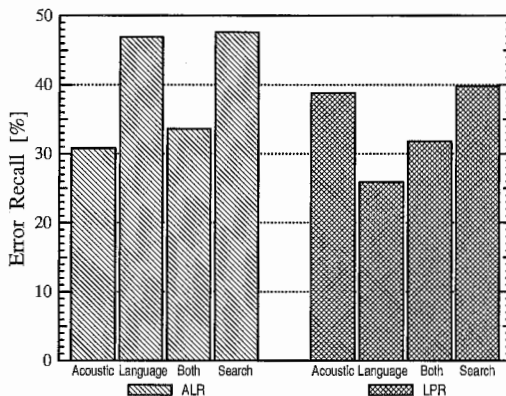


図 1.4: 誤り原因ごとの誤り再現率の比較

誤り再現率より高くなることが示された。これは、直感から考えられる傾向とは逆の傾向であるようであるが、次のように考えることができる。たとえば、観測された音声音が音響的にあいまいで音響モデルによる制約が十分に働かない場合、言語モデルの誤りを是正することができずに認識誤りを起こす。逆のケースもまた同様である。このように音響モデルと言語モデルによって候補を生成・選択する LVCSR のような音声認識システムでは、音声認識誤りは、音響的なあいまい性と言語的なあいまい性の力関係で起きることが多いと考えられる。したがって **ALR** や **LPR** といった信頼度尺度も、このバランスの崩れた部分で発生する認識誤りを多く検出したと考えられる。またこのことは、前節の実験で音響と言語の両方の要因による信頼度尺度（たとえば、競合単語仮説数など）を用いた場合に、最もよく認識誤りを検出できた事実とも一致する。

前節の実験で用いたその他の信頼度尺度についても、同様の実験を行なった。この結果を表 1.2 にまとめる。この表では、誤り原因ごとに比較して、再現率の高いものから順に © ○ △ × の記号によって、結果を示している。

この結果からも、上述した信頼度尺度の傾向は確認される。またその他の顕著な特徴として、探索エラー原因による誤りに対する再現率が高いことが挙げられる。探索エラー原因とは、ビームによる枝刈りによって、正解候補が刈られてしまうような場合に起きる誤りであり、局所的な音響スコアの落ち込みや非常にまれな言語表現などによって起きると考えられる。このような場合には、候補の中間表現である単語グラフに、正解候補が含まれない。このため信頼度尺度の計算を誤り語だけを用いて行なうこと

信頼度尺度	音響原因	言語原因	両原因	探索エラー
音響的尺度				
AL	○	×	△	◎
ALR	×	○	△	◎
NAL	×	◎	△	○
PH	△	◎	○	×
PD	×	○	△	◎
PHD	×	△	○	◎
WDL	×	△	◎	○
言語的尺度				
LP	○	△	×	◎
LPR	○	×	△	◎
CP	○	△	×	◎
WE	△	○	×	◎
BLP	◎	×	○	△
BCP	◎	×	○	△
BWE	○	△	×	◎
TP	△	×	○	◎
TS	○	△	×	◎
音響 + 言語 または 探索に関連する尺度				
TPR	×	△	○	◎
NSD	×	○	△	◎
NEM	×	○	△	◎
WH	△	×	○	◎
WPP	△	○	×	◎

表 1.2: 各尺度の誤り検出比較

になり、いずれの尺度を用いた場合にも信頼度が低くなったものと考えられる。

1.5 複数尺度の組合せによる認識誤り検出

これまで、単一尺度による認識誤りの検出について評価を行ってきた。これらの実験の結果も示すように、各尺度には検出できる誤りのタイプに偏りがある。このため、先行研究においても、これら複数の尺度を組合わせて用いるという手法が採られてきた。そこで本節では、効果的な尺度の組合せを得ることを目的とした、複数尺度による認識誤り検出の評価実験について述べる。

1.5.1 信頼度尺度の相関係数

これまで述べてきた各信頼度尺度は、認識誤りに関与する度合が異なると考えられる。複数の尺度を組合せる場合、関与の度合が高い尺度を合わせることで、より効果的に認識誤りが検出できるものと考えられる。そこでここでは、各信頼度尺度が認識誤りに関与する度合、すなわち認識誤りとの相関について調べる。

前節までの実験で行なった、信頼度尺度をそれぞれ単独に用いるような場合においては、認識誤りへの関与は、統計的には相関係数によって表わされると考えられる。しかし本節で行なうように、複数の信頼度尺度を組合せる場合には、互いに他の尺度の影響を除外した偏相関係数を用いることが妥当であると考えられる。そこで各信頼度尺度について、認識正解 (1) / 誤り (0) との相関係数と偏相関係数を表 1.3 にまとめる。なお偏相関係数については、計算不能になる“後向き単語出力カスコア (BWE)”を除外した 20 の尺度を用いて計算している。

この表に示したように、相関係数と偏相関係数では傾向にかなりの違いがある部分がある。たとえば“競合単語仮説数 (WH)”は、1.3 節の実験で最も良い検出性能を示しており、相関係数でも上位の値を示しているが、偏相関係数では下位の値となっており、他の尺度によって説明されうる尺度であることが示されている。逆に“音素持続時間確率 (PHD)”や“単語音素長 (WDL)”は、相関係数では低い値で

信頼度尺度	相関係数	偏相関係数
音響的尺度		
音響尤度 (AL)	0.111	-0.012
音響尤度比 (ALR)	0.243	0.027
正規化音響尤度 (NAL)	-0.212	-0.026
音素仮説数 (PH)	-0.159	0.016
音素距離 (PD)	-0.246	-0.056
音素持続時間確率 (PHD)	0.092	0.065
単語音素長 (WDL)	0.079	0.076
言語的尺度		
言語スコア (LP)	0.273	-0.031
言語スコア差 (LPR)	0.012	0.033
クラス遷移スコア (CP)	0.243	0.028
単語出力スコア (WE)	0.208	0.041
後向き言語スコア (BLP)	0.079	-0.002
後向きクラス遷移スコア (BCP)	0.093	-0.001
後向き単語出力スコア (BWE)	0.040	-
単語遷移パープレキシティ (TP)	-0.279	-0.081
トレーニングサンプル数 (TS)	0.247	0.070
音響 + 言語または探索に関連する尺度		
トータルスコア差 (TPR)	0.243	0.062
単語始端分岐数 (NSD)	-0.441	-0.072
単語終端分岐数 (NEM)	-0.440	0.042
競合単語仮説数 (WH)	-0.431	0.024
単語事後確率 (WPP)	0.520	0.245

表 1.3: 各尺度の相関係数

あり、単独で用いた 1.3 節の実験では極めて低い検出性能であったが、偏相関係数では比較上高めの値となっていて特徴的な要因を示す尺度となっていると考えられる。

1.5.2 3 尺度の組合せによる比較実験

以上のことを踏まえ、ここでは 3 つの信頼度尺度を組合わせる場合において、尺度の選択についての比較実験を行なう。

先行研究において、このように複数の尺度を合わせて用い認識誤りの検出を行なう場合には、ニューラルネットなどが多く用いられてきた [8]。ここではニューラルネットに合わせて、高い汎化性能を持つ学習機械として近年注目を集めている、SVM を用いた実験も行うものとする。実験では、3 つの尺度を用いた判別ルールを、ニューラルネット (MLP) および SVM で学習し、評価するものとする。評価尺度は 1.3 節の実験と同じで、認識誤りの再現率と適合率が等しくなる閾値で誤り再現率と判別率を比較する。ニューラルネットおよび SVM の学習には、これまで同様旅行対話 930 発話を用い、評価用データとしては、同じく ATR 旅行会話データベースから異なる 330 発話 (約 4700 語) のデータを用いた。3 つの尺度の選択基準には、1) 1.3 節の実験で、音響・言語・音響 + 言語の 3 つの各カテゴリごとの検出性能の上位 (ALR, LPR, WH)、2) 全尺度の中で検出性能の上位 (WH, WPP, NSD)、3) 偏相関係数の絶対値の上位 (WPP, TP, WDL)、の 3 通りについて比較する。この実験の結果を表 1.4 に示す。なお、比較のため単独の尺度 (競合単語仮説数 (WH)) を用いた結果も合わせて示す。(ただし単独尺度の場合は、単に閾値判定によって性能を評価している。)

ここに示された結果から、まず、単独の尺度を用いた検出より複数尺度を用いた方が性能が上がる事が確認された。また、複数尺度を組合せる場合には、用いる尺度の選択如何によって識別性能が変化することが確認された。今回の比較においては、全尺度の認識誤りに対する偏相関係数が、最も高いものを 3 つ選んだ場合が最良の結果となった。また、学習機械として MLP と SVM を用いた結果の比較においては、この両者について十分な差異があるとは言えない結果となった。ただし今回の評価においては、予備実験でいくつかの代表的な SVM カーネルを用いた比較をしたものの、最適な SVM カーネルについ

組み合わせ	MLP		SVM	
	ER [%]	DR [%]	ER [%]	DR [%]
1) ALR,LPR,WH	49.4	89.7	49.4	89.8
2) WH,WPP,NSD	53.1	90.5	53.4	90.5
3) WPP,TP,WDL	56.5	91.1	53.7	90.6
WH	50.9	89.1		

表 1.4: 3尺度を用いた認識誤り検出の比較

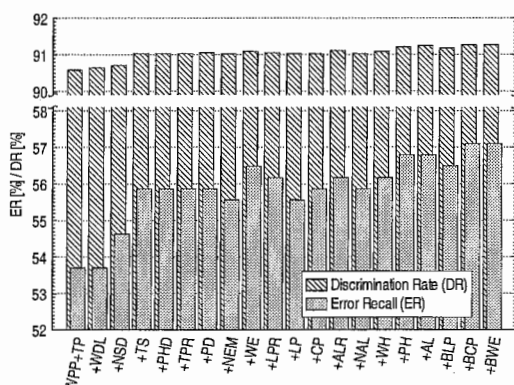


図 1.5: 多尺度による誤り検出

て十分な検討がなされたとは言えず、この点については検討の余地がある。

1.5.3 多尺度の組合せによる評価

前述の実験では、予備的に3つの尺度のみを用いた認識誤り検出の評価を行なった。これにより偏相関係数の高い尺度を組み合わせることで、高い検出性能が見込まれることを確認したが、さらにここでは、より多くの尺度を組合わせて検出性能の変化を調べる。すなわち、認識誤りの検出に用いる尺度として、1.5.1節に述べた偏相関係数の絶対値が大きいものを、“単語事後確率 (WPP)” から順に加えて、検出に利用する尺度の数を1つずつ増やして、検出性能の変化を調べるものとした。実験条件はこれまでと全く同様とする。評価尺度もこれまで同様、認識誤り再現率と認識誤り検出率が等しくなる閾値での、認識誤り再現率 (ER) と判別率 (DR) を用いる。複数尺度併用のための判別器としては、SVMを用いるものとした。これは、前節の実験でMLPとSVMの結果に大きな差がなかったこと、SVMは本質的に局所解に落ちこまない手法であると考えられているため、トレーニングセットの偏りに起因する識別性能の揺れが少いと期待できるからである。

このような条件の元での実験結果を、図1.5に示す。この結果から、認識誤りの検出に用いる尺度の種類を増やすにつれ、検出性能も向上することが確認された。また、再現率・判別率の変化から、尺度として5種類程度の尺度を組合わせたあたりで、性能の増加幅が小さくなっていると見ることができる。これは、尺度の選択基準として利用している、偏相関係数の絶対値の変化と比べて考えれば、妥当な傾向であると考えられることができる。したがって、全ての尺度を誤り検出に利用できない場合においても、偏相関係数を基準として選択した尺度を組み合わせることで、良好な性能を確保できることを示していると考えられることができる。

なお今回の評価実験では、21の尺度全てを用いた認識誤り検出による性能が最も高く、等再現率・検出率において誤り再現率57.1%、判別率91.3%を達成した。この値は“競合単語仮説数 (WH)”を単独で用いて誤り検出を行った場合に比べて、判別誤り率 (1-判別率) でおおよそ20%の改善を意味する値となっている。

1.6 まとめ

本論文では、認識誤りの有効な検出方法の確立を目的として、これまであまり検討されてこなかった種々の信頼度尺度による認識誤りの検出能力や検出傾向の違いについての検討を行なった。単一尺度による認識誤りの検出実験では、競合単語仮説数を尺度として用いる場合が最も良い結果となることを示した。また、音響的尺度あるいは言語的尺度だけによる推定では、十分な検出精度が得られないことを示した。

また認識誤りの検出傾向の比較実験から、音響的尺度の信頼度尺度は、主に音響モデルに起因する認識誤りより、主に言語モデルに起因する認識誤りの方をより多く捉えていること、逆に言語的尺度の信頼度尺度は、主に音響モデルに起因する認識誤りの方をより多く捉えていることを示した。このことは、音響モデルと言語モデルの2つのモデルの制約によって解を求める音声認識システムにおいて、いずれか一方ではなく、音響・言語両者の関係を捉える尺度が認識誤りを検出する上で有効であることを示唆するものである。

さらに、複数尺度を組合わせて認識誤りを検出する場合においては、統計的な立場から、偏相関係数に基づいた尺度の選択が有効であることを示した。より多くの尺度を組合わせた場合には、より高い認識誤り検出の性能が得られることを示したが、偏相関係数の絶対値の大きな尺度を数個利用するだけでも、比較的良好な検出性能が得られることを示した。

参考文献

- [1] T. Kemp et al.: "Estimating Confidence Using Word Lattice", In Proc. Eurospeech'97, pp. 827-830, 1997
- [2] F. Wessel et al.: "Using Word Probabilities as Confidence Measures", In Proc. ICASSP'98, pp. 225-228, 1998
- [3] E. Lleida et al.: "Efficient Decoding And Training Procedures For Utterance Verification in Continuous Speech Recognition", In Proc. ICASSP'96, Vol.1, pp. 507-510, 1996
- [4] 駒谷 和範ら: "混合主導対話における音声認識誤りに対処するための対話管理", 言語処理学会第6回年次大会, pp. 336-339, 2000
- [5] R. El méliani et al.: "Accurate Keyword Spotting Using Strictly Lexical Filler", In Proc. ICASSP'97, Vol.2, pp. 907-910, 1997
- [6] D. Klakow et al.: "OOV-detection in Large Vocabulary System Using Automatically Defined Word-Fragments As Fillers", In Proc. Eurospeech'99, pp. 49-52, 1999
- [7] R.A. Sukkar et al.: "Vocabulary Independent Discriminative Utterance Verification for Non-keyword in Subword Based Speech Recognition", IEEE Trans. on Speech and Audio Processing, Vol.4, No.6, pp. 420-429, 1996
- [8] T. Schaaf et al.: "Confidence Measures for Spontaneous Speech Recognition", Proc. of ICASSP'97, Vol 2, pp. 875-878, 1997
- [9] S. Cox et al.: "Confidence Measures for the Switchboard Database", In Proc. ICASSP'96, pp. 511-514, 1996
- [10] B. Rueber: "Obtaining Confidence Measures from Sentence Probabilities", In Proc. Eurospeech'97, Vol.2, pp. 739-742, 1997
- [11] A.R. Setlur et al.: "Correcting Recognition Errors via Discriminative Utterance Verification", In Proc. ICSLP'96, Vol.2, pp. 602-605, 1996
- [12] 山本 博史ら: "接続の方向性を考慮した多重クラス N-gram 言語モデル", 信学論, D-II, No.11, pp. 2146-2151, 2000.11
- [13] A. Nakamura et al.: "Japanese Speech Databases for Robust Speech Recognition", In Proc. ICSLP'96, pp. 2199-2202, 1996
- [14] Morimoto et al.: "Speech and Language Database for Speech Translation Research", In Proc. ICSLP'94, pp. 1791-1794, 1994
- [15] G. Evermann et al.: "Large Vocabulary Decoding and Confidence Estimation Using Word Posterior Probabilities", In Proc. ICASSP'00, Vol.3, pp. 1655-1658, 2000
- [16] 河原 達也: "大語彙連続音声認識における認識誤り原因の自動同定", 音響学会講演論文集, 2-1-17, pp. 87-88, 1999.9

第 2 章

旅行会話基本表現コーパスを用いた認識誤り訂正の検討

2.1 はじめに

近年多くの音声認識システムが開発され、その幾つかは実用レベルに近付きつつある。しかし、現状の認識システムにおいて誤認識の問題は避けられない。従来は認識誤りを、モデルと探索手法の改良によって低減するアプローチが採られてきたが、我々は認識誤り箇所の検出とこれの訂正というアプローチの検討を進めている。これは、従来の音声認識の枠組によって得られる認識結果について、より広い範囲のコンテキスト情報などを用いることで、局所的な認識誤りを訂正できるとの考えに基づく。

我々は過去において認識誤り箇所の検出について検討を行ってきた [1]。今回は認識誤り箇所の訂正方法について検討を行なう。認識誤り箇所訂正のための先行研究は、OCRの分野 [2, 3] で多くなされてきた。しかし、文字認識に比べて、音声認識では入力音声の曖昧性をはるかに大きいため探索空間が広がり過ぎて、これらの手法を音声認識にそのまま適用するのは難しい。音声認識の分野では、例えば正しく認識できた部分のみ翻訳する方法 [4] などが提案されているが、誤り訂正そのものを行なう検討はあまり多くなされていない。

その中で、石川ら [5] は、検出された誤り箇所に対して音韻的に類似した用例を用いて代替候補を生成し、音韻的距離と意味的距離によって候補の妥当性を判断する誤り訂正法を提案して一定の成果を収めた。しかしこの方法で用いている意味距離は、多義語の問題などのために十分な制約力を持たず、またオープンテストに対する性能も良くなかった。

我々が対象としている旅行会話のような対話における多くの発話は、会話表現集に見られるように単文を中心とした比較的単純な構造であり、大量の用例を収集することによって頻度の高い表現をカバーすることができると考えられる。そこで我々は、用例コーパス中の用例文を用いて、用例文そのものを制約とする誤り訂正法を提案する。この提案法では、認識結果文の類似用例を用例コーパスから探索し、これを基に誤り箇所の代替候補を生成し訂正を行なう。評価実験の結果、提案方法によって 20% 以上の誤り箇所が正しく訂正できることが確認された。

2.2 提案方式

音声認識器は通常複数の認識候補 (N-best) を出力するが、本論文ではこのうちの 1 位候補 (1-best) の系列のみを用いて、その誤り箇所を訂正する。この認識誤り箇所訂正の問題をここでは次の 3 つに分けて考える。

1. 認識誤り箇所の検出
2. 代替候補の生成
3. 最適候補の選択

この中で認識誤り箇所の検出の問題は重要なテーマであるが、他稿に譲り ([8, 9] など)、本論文では後者 2 つについて検討を行なう。

我々は認識誤り箇所訂正のための情報源として、認識結果文と類似する用例文そのものを用いることを提案する。これは、同一ドメインで同じ内容を伝える文はいくつかの類似した文に大別されるであろうという仮定に基づいており、また用例文という強い言語的制約によって、誤り箇所が訂正されることを期

文数	92,668
平均文長	7.4
総単語数	681,525
語彙数	14,419

表 2.1: 用例コーパス

待するものである。従って本提案手法は、あらゆる発話の認識誤りを保証するものではなく、認識対象と同一ドメインの用例コーパスによって規定された、対象領域の発話の認識誤りを訂正するものである。

2.2.1 本研究における認識誤り箇所

本論文においては、問題を明確にするため、認識誤り箇所は正確に検出できるものとして検討を行なう。すなわち、あらかじめ手作業で与えた正解の単語ラベルと、認識器の出力する 1-best の単語系列の単語ラベルを比べて、置換および挿入誤りであった部分を認識誤り箇所としている。脱落誤り箇所については、認識結果中に該当単語が表われないため検出できないので無視する。

2.2.2 代替候補の生成

検出された音声認識誤りの箇所に対する代替候補を生成する。代替候補の探索範囲として用例コーパスを用いる。すなわち認識結果文と類似する用例をコーパス中から探索し、これに表われる単語を当てはめる。類似用例の選択には、コーパス中の各文と認識結果文で DP マッチングに基づく距離計算を行ない [6]、これを類似度として順序付けを行なう。またこれと同時に得られる、認識結果文の各単語と用例文の各単語の対応関係から、認識結果の誤り箇所に対応する単語を用例文より取り出し代替候補とする。

用例コーパス

我々は、旅行会話音声翻訳器のための音声認識部の開発を進めている [7]。したがって我々の音声認識器の認識タスクは、旅行会話音声である。このタスクの音声の認識誤りを訂正するための用例コーパスとして、旅行者向けのフレーズブックに表われるような旅行会話の基本表現を大量に集めたものを用いる。集められた用例文の数は、重なりなしでおよそ 9 万文である。表 2.1 にこの用例コーパスの統計値をまとめる。

用例の選択

認識誤り箇所の代替候補の集合を生成するため、用例コーパスから認識結果文の類似文を検索する。類似文の探索は認識結果文と各用例文の DP マッチングに基づいて行なう。認識結果文中の認識誤り箇所については、“誤り単語”を意味する特別な ID を与えて、他のいかなる単語とも異なる単語としてマッチングを行なう。認識結果文と用例文の距離を定義する式を以下に示す。

$$dist = \frac{I + D + \sum d_i}{L_{in}} \quad (2.1)$$

ここで I および D は、それぞれ挿入誤りと脱落誤りの回数を示しており、 d_i は置換誤りにおける単語間の距離を示している。また L_{in} は、認識結果文の単語系列長を意味する。

単語間の距離としては、いくつかの方法を考えることができる。最も単純には、 $d_i = 1.0$ という定数値を与える方法である。今回はこのような単語間距離に加えて、単語間の意味距離を用いる方法についても検討した。これは、シソーラス上での意味属性の位置関係により単語間に 0~1 の意味距離を与えるものである (詳細は後述)。

以上のような距離計算によって各用例ごとに得られた類似度のうち、上位 n 位の類似度を有する用例を代替候補単語を含む文として選択する。

2.2.3 最適候補の選択

認識誤り箇所に対する代替候補の集合中から、誤り箇所に最も当てはまると考えられる単語を選択する。これには認識結果の当該箇所の音素系列と、発音辞書に示された代替候補単語の音素系列を DP マッ

音響モデル	HMnet (コンテキスト依存, 男女別)
言語モデル	多重クラス複合 2-gram
探索	2パス, ビーム探索
辞書	語彙数 38,173 語

表 2.2: 音声認識システムの概要

チングさせた結果を用いて, 次に示すスコアによって行う.

$$score = \lambda \cdot dist + (1 - \lambda) phdist \quad (2.2)$$

$$phdist = \frac{I_{ph} + D_{ph} + S_{ph}}{L_{ph}} \quad (2.3)$$

ここで I_{ph} D_{ph} S_{ph} は, それぞれ認識結果の音素系列と, 代替候補単語の音素系列の DP マッチングによる挿入誤り, 脱落誤り, 置換誤りの回数である. また L_{ph} は, 誤り区間の音素系列の系列長である.

この式は, 認識結果文と用例文の文としての近さと, 単語の音素系列の近さ—すなわち音響的近さ, の2つの要因の重み付け和によって代替候補単語を順序付けすることを表わす.

2.3 提案方式の評価

2.3.1 実験条件

提案方式の基本性能を評価する実験を行なう. 実験に用いた音声認識システムの概要を表 2.2 にまとめる.

認識に用いた音声は, 2.2.2 節で述べたコーパスを作成する際には利用しなかった別の旅行会話文セットを, 複数の男女が発声したものを, 総発話数は 2,037 発話であった. 前述の音声認識システムによる評価では, この音声データに対する認識精度は単語 Accuracy で 83.9% であった. 正解の単語ラベルと認識結果を比較した結果では, 計 897 箇所認識誤りがあった.

実験では, 2.2.2 に述べた用例セットを用いて認識誤り箇所の訂正を行なう. 用例の類似度 (文間距離) の計算は式 (2.1) に基づいて行ない, 単語置換の距離 d_i は一律に 1.0 を与えた. 代替候補の順序付け式 (2.2) では, 認識結果の音素系列として, 誤り区間の 1-best の単語の音素系列をそのまま用いるものとした.

2.3.2 実験結果

以上に述べたような条件の下で, 次の点に着目した評価を行なった.

- 1つの誤り箇所に対する平均代替候補数
- 全誤り箇所に対して, 正解候補を含む代替候補が得られた誤りの箇所の割合 (可能誤り訂正率, correctable error rate)
- 全誤り箇所に対して, 代替候補中から正しく正解候補を選び出すことのできた誤り箇所の割合 (誤り訂正率, corrected error rate)

ただし認識誤り箇所とは, 連続した誤り単語をまとめた区間を指す. これは, 本提案方式では連続して誤っている区間全体をまとめて訂正を行なうためである.

この結果を図 2.1, 2.2 に示す. 図 2.1 は, 類似度 (文間距離) の上位何位までを類似用例とみなすか (上位選択順位数, N-best) に対する平均代替候補数および可能誤り訂正率を示すものである. 上位選択順位数に対する平均代替候補数は, 縦軸右側のスケールで示されており, 可能誤り訂正率は縦軸左側のスケールで示されている.

本来, 上位選択順位数と平均代替候補数は近い値となると予想されるが, ここでは平均代替候補数が非常に多い. これは, 同じ類似度を持つ用例が非常に多く表われるケースがあるためである (たとえば, 「<Error>をお願いします」という誤り文では, 非常に多くの類似用例が得られる). 図 2.1 に示された結果からは, 上位選択順位数を大きくするに伴ない, より多くの認識誤り箇所に対して正解候補を含む代替候補が得られるようになるが, 選択順位を 10 位以上に広げても, 可能誤り訂正率は伸びないことが示され

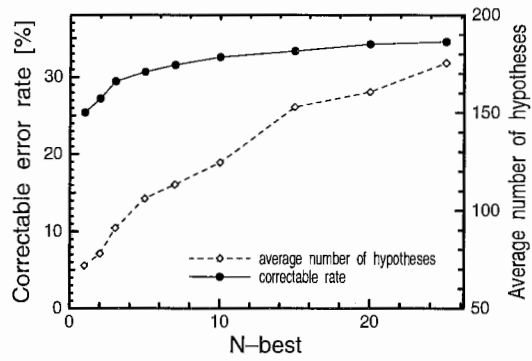


図 2.1: Average number of hypotheses & Correctable error rate

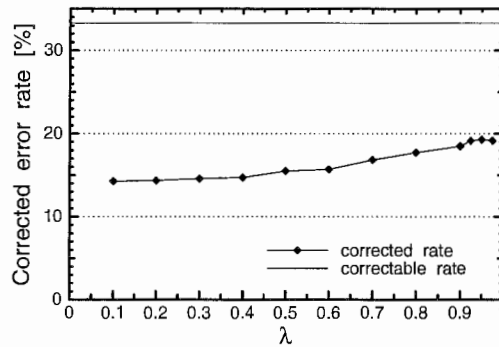


図 2.2: Corrected error rate

ている。また選択上位順位数を 15 位とした場合で、可能誤り訂正率は約 33% であり、本実験で行なった方法で、1/3 の誤り箇所が正しく訂正できる可能性があることを示している。

また図 2.2 は、代替候補を 2.2.3 節に述べた方法によって順序付けた場合の誤り訂正率を示している。このグラフにおいて横軸は、式 (2.2) における λ の値を示し、縦軸は誤り訂正率を示している。このグラフでは、上位選択順位数は 15 位としている。比較のため、図 2.1 に示した上位選択順位数 15 位における可能誤り訂正率を、誤り訂正率の上限値として合わせて示す。この結果から、本手法により全誤り箇所の約 20% を正しく訂正できることが示されている。

2.4 音素距離の検討

前節の実験では、用例により得られた代替候補を式 (2.2) に基づいて順序付けする際、1 位の (誤り) 認識単語の音素系列と、代替候補単語の音素系列の類似度を用いた。しかし音声認識の過程で、単語候補の生成には音響モデルと言語モデルという 2 つの制約が用いられている。このため、誤り認識単語の音素系列は言語的な制約による影響を受けていおり、しかもこれは誤った影響である可能性がある。

そこで本節では、音響モデルだけの制約によって生成される音素系列を用いた場合についての比較実験を行なう。具体的には、音声認識の過程でモーラタイプライタを並行に走らせ、認識誤り区間と判定された区間に該当する音素系列を、このモーラタイプライタ結果から取り出す。

この実験の結果を、先の誤り認識単語の音素系列を用いた場合と合わせて図 2.3 に示す。縦・横軸の意味は図 2.2 と同じで、音素類似度と文類似度の重みに対する、誤り訂正率の関係を示している。この図に示されるように、参照する音素系列としてモーラタイプライタによるものを用いた方が、より多くの誤り箇所を正しく訂正できることが示されている。

この結果に基づき、以下の評価実験では最適候補選択のための音素系列として、モーラタイプライタによる音素系列を用いるものとする。

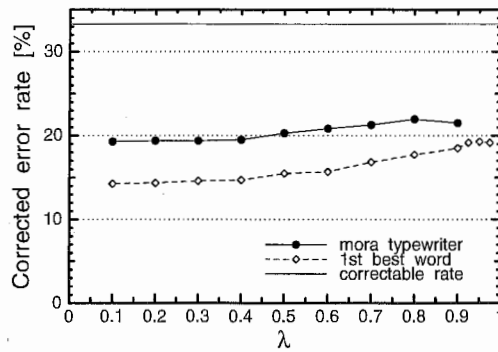


図 2.3: Rescoring by using mora typewriter

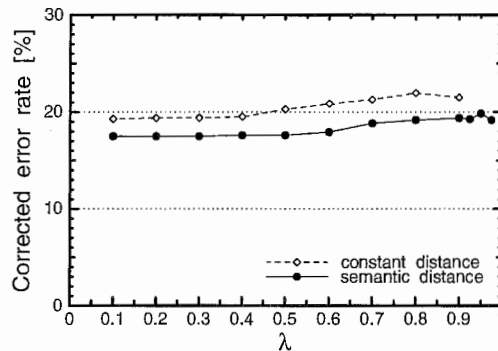


図 2.4: Searching examples by using semantic distance

2.5 文類似度の評価

2.5.1 文類似度

これまでの実験では、認識結果文と用例文の距離として、単語間の距離を定数値 1.0 とした結果について示した。本節では、単語間の距離を単語ごとに変えた場合について評価する。

本節で行なう実験では、単語間の距離を意味距離に基づいて可変とする。ここで言う意味距離は、シソーラス上での単語意味属性の一致度で計算をする。これは単語のシソーラス階層数が N で、2つの単語の意味属性が最上層から第 $N - k$ 階層まで共通であるなら、以下の式によって計算される。

$$d_{sem} = \frac{k}{N} \quad (2.4)$$

今回の実験では、シソーラスの階層数は $N = 3$ で固定としているため、 $d_{sem} = \{\delta, \frac{1}{3}, \frac{2}{3}, 1\}$ の 4 種類の値を取る。(d_{sem} の最小値を δ としたのは、同一語間の距離を 0.0 とし、これとの差別化を図るためである。)

しかし、通常 1 つの単語には複数の語義が付される。このため上式のみでは、単語間の距離を決定することができない。本実験では、以下の式により評価を行なった。

$$d(w_1, w_2) = \min_{i,j} d_{sem}(w_{1i}, w_{2j}) \quad (2.5)$$

ここで w_{ni} は単語 n の第 i 番目の語義を示す。この式は、単語の複数語義の間の最小意味距離を当該単語間の意味距離とすることを意味する。

2.5.2 実験結果

実験結果を、単語間距離を 1.0 に固定した場合と比較して、図 2.4 に示す。

この結果からは、用例検索において単語間の意味距離を用いることによる効果は表われていない。この原因は次のように考えられる。

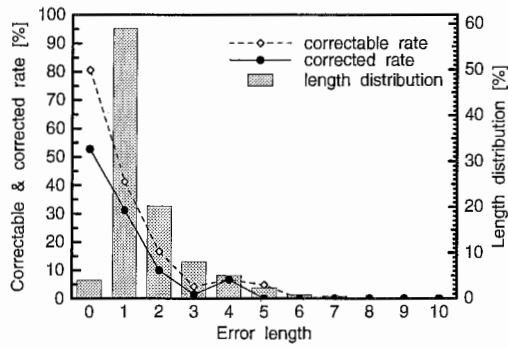


図 2.5: Error length distribution

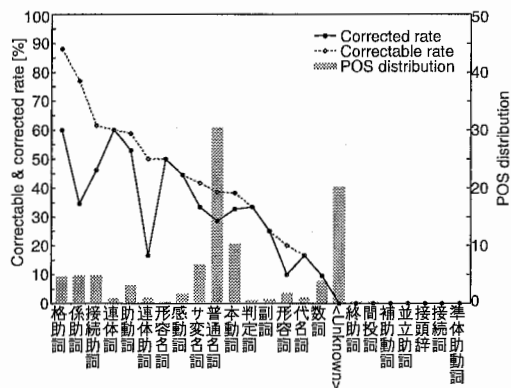


図 2.6: Error POS distribution

認識結果文と用例文を対応させて文類似度を計算する際、認識結果中の認識誤り箇所は、“誤り単語”を意味するIDによってマスクされており、いかなる単語とも距離は1.0となる。このため肝心な箇所の意味距離が働かず、その効果が低減してしまっている。

また、単語が持つ複数の語義を解消する方法として、今回、最小距離を用いたが、このような単純な方法では十分な解消ができていないことも原因として挙げられる。

2.6 誤り訂正結果の分析

2.6.1 誤り長ごとの評価

本論文では、誤り箇所とは連続する誤り単語の区間を指す。本提案方式によって、どの程度の誤り長（誤り区間における正解単語の系列長）の誤りを訂正できるのか調査した。

この結果を誤り長に対する可能誤り訂正率、および誤り訂正率の関係として図2.5に示す。また同時に、各誤り長の誤りが、全誤りの中でどの程度の割合で出現するかを棒グラフによって重ねて示している。本図において、可能誤り訂正率および誤り訂正率は、縦軸左側のスケールで示されており、各誤り長の割合は縦軸右側のスケールで示されている。

この実験の結果から、本提案方式により訂正可能な誤りは誤り長2程度までの誤りであって、これ以上の長い誤りについてはほとんど訂正の見込みがないことが示されている。なお誤り長2までの誤りは、全誤り箇所のおよそ80%であることもグラフから読み取れる。

2.6.2 品詞ごとの評価

誤り箇所の単語に対し、その品詞ごとの可能誤り訂正率、誤り訂正率を比較する。ここでは、長さ1の誤り箇所のみについて、その正解単語の品詞ごとに分類する。

この結果を図 2.6 に示す。図中で棒グラフは、誤り長 1 の全ての誤り箇所に対して、各品詞の占める割合を縦軸右側のスケールで示している。また各品詞ごとの可能誤り訂正率および誤り訂正率を、折れ線グラフによって縦軸左側のスケールで示している。

この結果からは、第 1 に、格助詞といった機能語の誤りについてはこれらの語が前後のコンテキストから比較的容易に正解語が類推できること、内容語に比べて種類が少ないこと等の理由により、比較的高い割合で代替候補中に正解語が含まれることが判る。しかし、最適候補選択においては、正解候補を正しく選択できていない。これは、これらの語が 1, 2 音節程度の短かい語であるため、候補選択において音素距離による差がつきにくいいためであると考えられる。

また第 2 に、普通名詞、本動詞、形容詞といった内容語については、代替候補を選択した段階ですでに正解語が含まれていないことが多い。これは、これらの品詞は単語の種類が多いため、同様のコンテキストで交換可能な語が多く、用例文から得た代替候補のみでは候補生成が不十分であるためと考えられる。しかし、棒グラフで示された誤りに占める品詞の割合から言っても、また文の内容を正しく捉えるという意味においても、特に普通名詞、本動詞の誤りは、訂正の効果を上げる上で重要なターゲットであると考えられる。

2.7 まとめ

本論文では、音声認識の誤り箇所の訂正方法として、用例コーパスを用いて、認識結果文の類似用例を検索し、これによって誤り箇所を訂正する方法を提案した。

用例の検索方法、代替候補のリスコアリング方法などいくつかの比較実験を行なった結果、全認識誤り箇所のうちの 22.0% を正しく訂正できることを確認した。また、リスコアリングにより正しく正解候補を導くことのできなかつた誤り箇所でも、代替候補中には正解候補が含まれているケースがあり、適切なりスコアリング方法を用いることでさらに訂正性能が向上する可能性がある。

また実際に訂正できた誤り箇所についての分析から、本訂正手法は、全誤りの 80% を占める、誤り系列長が 2 以下の誤りに対して効果的であることが確認された。さらに、品詞ごとの訂正性能の比較から、多く誤っている普通名詞や本動詞といった品詞の誤り訂正が不十分であることが示された。これは、単語の共起関係を用いるなど、このターゲットに絞ったアプローチによってさらに訂正性能を向上させることが可能ではないかと考えている。

参考文献

- [1] Y.Okimoto et al.:“Evaluation of Mis-Recognition Detection Using COnfidence Measures”,In Proc. ICSP’01, pp. 685–690, 2001
- [2] 竹内孔一ら:“統計的言語モデルを用いた OCR 誤り訂正システムの構築”, 情報処理学会論文誌, 40 (6), pp.2679–2689
- [3] 永田昌明:“文字類似度と統計的言語モデルを用いた日本語文字認識誤り訂正手法”, 電子情報通信学会論文誌, J81-D-II (11), pp.2624–2634
- [4] 脇田由美ら:“意味的類似性を用いた音声認識正解部分の特定法と正解部分のみ翻訳する音声翻訳手法”, 自然言語処理, 5 (4), pp.111–125, 1998
- [5] 石川開ら:“テキストデータを使った音声認識誤りの訂正”, 自然言語処理, 7 (4), pp.205–227, 2000
- [6] E.Sumita:“Example-based machine translation using DP-matching between word sequences”, In Proc. ACL-2001 Workshop (DDMT), pp. 1–8, 2001
- [7] T. Takezawa et al.:“A Japanese-to-English Speech Translation System: ATR-MATRIX”, In Proc. IC-SLP’98, pp. 957–960, 1998
- [8] T. Kemp et al.:“Estimating Confidence Using Word Lattice”, In Proc. Eurospeech’97, pp. 827–830, 1997
- [9] F. Wessel et al.:“Using Word Probabilities as Confidence Measures”, In Proc. ICASSP’98, pp. 225–228, 1998

付録 A

音声認識誤りの検出・訂正ツールについて

A.1 はじめに

このドキュメントでは、音声認識誤りの検出と訂正に関連したいくつかのツールの利用方法について解説する。音声認識誤りの検出問題と認識誤り訂正の問題に関しては、これまでそれぞれ個別に検討してきたため、ここに述べるツール群も、以下に示すように異なるツールセットとして構成されている。しかしながら、各々のツールセットが出力する各中間ファイルの形式は、共通のフォーマットに従っているため、認識誤り検出から誤り訂正への(処理上の)連結は容易であると考えられる。

音声認識誤りの検出に関するツール

ATR-SPRECの認識結果である単語ラティスファイルおよびN-bestファイルを入力として、単語信頼度尺度 (Confidence Measure, CM) の計算を行ったり、これを元に認識結果単語の正解・誤りの推定を行ったりするツールなどが含まれる。

音声認識誤りの訂正に関するツール

音声認識誤り箇所に関するラベルが付された認識結果文に対して、この文の類似用例文を検索したり、認識誤りとされた単語に対する代替候補を検索したりするツールなどが含まれる。

A.2 音声認識誤り検出のフロー

本節では音声認識誤り検出に関する処理フローについて解説する。この処理に関するデータフローを図 A.1に示す。なお、この図を含む本解説書中の処理フロー図では、実線の矢印は実線矢印で連結された全てのデータが同時に入力あるいは出力されることを意味している。また破線の矢印は破線の矢印で連結されたデータのうちのいずれか1つが選択的に入力あるいは出力されることを意味している。さらに、太線の矩形で示された部分が1つの処理を意味しており、本書で解説するいずれか1つのコマンドに対応している。

A.2.1 信頼度尺度計算のための前準備

信頼度尺度を計算するスクリプト **CalcCP.pl** は、標準の ATR-SPREC が出力する単語ラティスファイル・N-best ファイルでは処理ができない。**CalcCP.pl** が動作するためには、単語ラティスファイル・

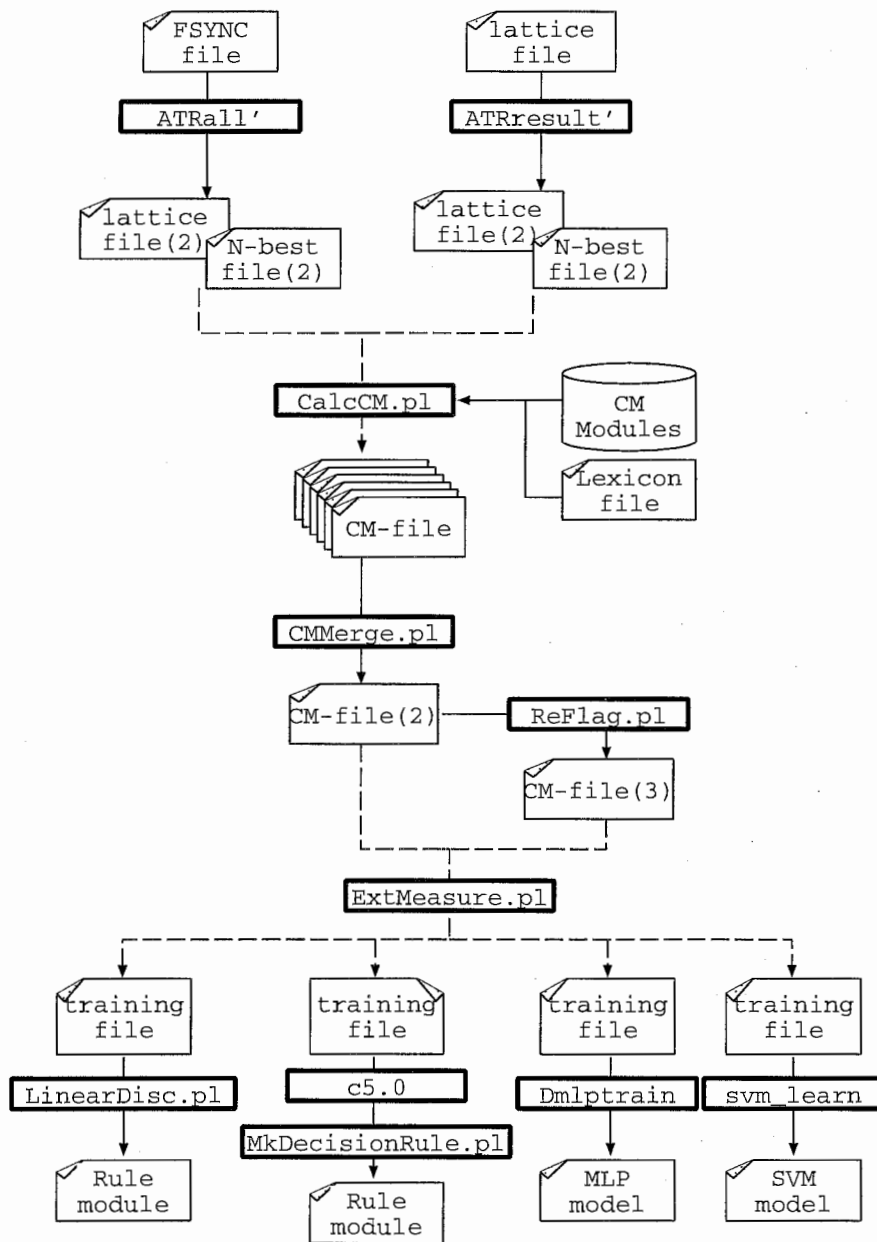


図 A.1: 認識誤り検出ツールデータフロー (トレーニング)

N-best ファイルに幾つかの追加情報を加える必要がある。

この目的のため、追加情報を出力することができるように改造された ATR-SPREC コマンドを用意した。これらを以下に示す。

- **ATRall**
FSYNC ファイルから単語ラティスファイル (lattice file) と N-best ファイル (N-best file) を生成
- **ATRresult**
単語ラティスファイル (lattice file) から N-best ファイル (N-best file) を生成
- **ATRtrace**
FSYNC ファイルから正解系列の単語ラティスファイル (lattice file) と N-best ファイル (N-best file) を生成

これらは ATR-SPREC(Ver.r07r03) をベースに改造したもので、基本的な機能・利用方法などはオリジナルのものと同じである。しかしこれらのコマンドでは、追加情報を出力させるために幾つかの実行オプションの追加、追加情報のための出力フォーマットの変更がなされている。追加出力される情報としては次のものが挙げられる。

- **lattice_arc**
N-best ファイルに追加出力される。この情報は、N-best ファイル中の当該順位の認識結果文に対応する単語ラティス中のパスを表わしている。N-best ファイル中の **lattice_arc** キーにリストされた数値は、単語ラティスファイル中の対応する単語ラティスのアーク ID のリストであり、このリストを辿ることでラティス中でのパスが示されている。この情報は **CalcCM.pl** の実行に必須情報であり、**ATRall**, **ATRresult**, **ATRtrace** 全てのコマンドでオプションなしで、出力される。
- **forward, backward**
単語ラティスファイルに追加出力される。単語信頼度尺度として単語事後確率に基づいた値を利用する場合に必要な情報である。このキーに続いて出力される数値は、当該ノードから見た前向き確率(対数確率)と後向き確率(対数確率)の値である。この追加出力を得るためには、**ATRall**, **ATRresult**, **ATRtrace** の実行時に次のような追加のオプションが必要である。**-ATRresult:FBProb=(1|2|3)** オプションの右辺は 1~3 のいずれかの数値を与えればよいことを意味しており、この値が大きい程より細かい情報が出力されることを意味する。(1: 音響+言語トータル, 2: 音響・言語別, 3: 音響・クラス遷移・単語出力スコア別)
- **ngram_info**
単語ラティスファイルに出力される。単語信頼度尺度としてクラス遷移確率、およびクラス出力確率に基づいた値を利用する場合に必要な情報である。このキーに続いて出力される数値は、クラス遷移確率、単語出力確率(対数確率)を意味する値で、この両者の和が理論上言語スコアに等しい。(計算の都合上、多少の丸め誤差が存在する。)この追加出力を行なうのは、**ATRall**, **ATRtrace** のみで、これらに対して **-ATRlattice:DivLangScore=yes** なるオプションを与える必要がある。

以上を表 A.1 にまとめる。

A.2.2 信頼度尺度計算

前節に挙げたコマンドによって生成した、単語ラティスファイルと N-best ファイルの組を入力データとすることによって、**CalcCM.pl** が信頼度尺度の計算を行なう。

CalcCM.pl では、種々の信頼度尺度の計算を容易に切り換えて計算することができるように、信頼度尺度計算の中心部分をモジュール化している。つまり、信頼度尺度のカテゴリごとに異なる信頼度尺度計算モジュールを、計算の都度取り換えてインポートすることによって信頼度尺度計算を行なうよう実装されている。このモジュールの取り換えのためには **CalcCM.pl** の冒頭部分でコメントアウトされている、複数の CM 計算モジュールの中から 1 つだけ必要なモジュールを選択、有効にして **CalcCM.pl** を利用

表 A.1: 単語ラティス, N-best 追加情報

追加情報キー	用途	対応コマンド	コマンドラインオプション
lattice_arc (N-best file)	各 N-best 文の単語ラティス上でのパス (CalcCM.pl 必須)	ATRall ATRresult ATRtrace	なし (常に出力)
forward backward (lattice file)	単語事後確率の計算	ATRall ATRresult ATRtrace	-ATRresult:FBProb=1,2,3
ngram_info (lattice file)	クラス遷移、単語出力スコアの計算	ATRall ATRtrace	-ATRLattice:DivLangScore=yes

する。各 CM 計算モジュールの詳細については A.5 節で詳述するが、各モジュールファイルそのものの冒頭コメントにも解説が書かれているので、こちらも参照されたい。

A.2.3 複数の信頼度尺度のマージ

CalcCM.pl は、1 度に 1 つの CM 計算モジュールを用いて、特定カテゴリの信頼度尺度を計算する。したがって、複数カテゴリにまたがる信頼度尺度を併用して機械学習を行ない、認識誤りの検出を行なうためには、これらの尺度をマージして 1 ファイルにまとめる必要がある。

CMMerge.pl は、異なるファイルに格納された信頼度尺度群を選択的に組み合わせて、1 つの信頼度尺度ファイルを生成する目的で利用する。

A.2.4 評価対象外単語の特定

CalcCM.pl や CMMerge.pl によって出力されたファイルを、学習機械でトレーニングすることができる種々のデータフォーマットに変換する前に、特定の条件の単語をトレーニング・評価対象から除外することを意味するフラグを付与することができる。(例えば間投詞の無視など。)

ReFlag.pl は、このような目的で利用する。ただし、もし除外単語がなくフラグ付けが不要であればこのコマンドによる処理はしなくてもよい。

A.2.5 学習機械用データの生成

以上に述べたコマンドによって生成されたファイルを整形して、何種類かの学習機械をトレーニングするためのデータを生成する。

これを行なうコマンドは ExtMeasure.pl で、線型判別、決定木、MLP(ニューラルネット)、SVM (Support Vector Machine) をトレーニングするためのデータを出力することが可能である。いずれの学習機械のトレーニングを行なうのかをコマンドラインオプションで与えることで、それに応じたファイル (場合により複数ファイル) を出力する。また、トレーニングされた各学習機械の判別性能の評価のためのテストセットデータも、同様の方法で作ることができる。

A.2.6 学習機械のトレーニング

学習機械のトレーニングは、各学習機械のプログラムのマニュアルに従って行なう。(線型判別だけは自前のツールであるので、後述する。) トレーニングによって得られる判別ルールモジュール、またはモデルファイルは、後述する誤り検出コマンドで利用される。

MkDecisionRule.pl は、決定木学習・評価プログラム (c5.0) のための補助ツールである。c5.0 は、-r オプションを与えてやることで学習した決定木 (判別ルール) を生成するが、そのままでは、誤り検出コマンドで利用することができない。誤り検出コマンドで利用可能なように、決定木を perl モジュールのフォーマットにするため MkDecisionRule.pl を用いる。

また MLP トレーニングコマンドである **Dmlptrain** は、トレーニングデータの他に project ファイル、データのリストファイルなどを必要とする。これらのサンプルファイルを自動生成する目的で、**MkMLPProj.pl** を利用できる。**MkMLPProj.pl** は、**ExtMeasure.pl** が生成したファイルを読み込んで、このデータに見合ったファイルを生成する。ただしこれによって生成されるファイルは、最低限の動作をするように書かれているだけであるため、適宜変更が必要である。

A.2.7 認識誤り検出

トレーニングされた学習機械を用いて、認識結果の各単語ごとの認識正解・誤りを推定させることができる。このためのコマンドは **ErrorDetect.pl** である。

ErrorDetect.pl は、前節で説明した学習機械のトレーニングにより得られた perl モジュールまたは、モデルファイルを元に、各認識単語ごとの正解・誤りを推定する。入力データファイルは、**CalcCM.pl**、**CMMerge.pl**、**ReFlag.pl** のいずれかによって生成されたファイルでなければならず、ファイルに格納されている信頼度尺度の構成は、学習機械のトレーニングの際の構成と同じでなければならない。

認識誤り検出に関するデータフローを図 A.2 に示す。

A.2.8 生成ファイルのフォーマット

本節では、認識誤り検出関連のツールが生成するファイルのフォーマットについて解説する。

CM-file

CalcCM.pl、**CMMerge.pl**、**ReFlag.pl** の各コマンドは、各認識結果の単語ごとの信頼度尺度値を出力する。出力ファイルは、1 行につき 1 単語の情報が出力される。各行に含まれる情報は、正解単語とこれに対応する認識結果単語、正誤フラグ、および 1 つ以上の単語信頼度尺度のリストなどである。この例を図 A.3 に示す。この例に示すように、各発話データの開始位置には、まず発話番号を表わす行が示され、それ以降に発話中の各単語ごとの情報が示される。単語情報の各行でのカラムの意味を表 A.2 にまとめる。第 6 カラム以降の各カラムの値の意味は、信頼度尺度の取得方法 (**CalcCM.pl** 実行時にインポートされた信頼度尺度計算モジュール) に依存する。これについては、A.5 節に後述する。(もしくは、信頼度尺度計算モジュールを直接参照されたし。) また先程の例では、単語情報行の中に、信頼度尺度の全てのカラムが“-”となっているものがあるが、これは当該単語が連結語の一部であるために、個別に信頼度尺度を求めることができないことを意味する。

Eval-file

認識誤り推定ツール **ErrorDetect.pl** は、認識結果の各単語の正解・誤りを推定した結果を、推定結果ファイル (Eval-file) に出力する。このファイルの例を図 A.4 に示す。ここに示すように、認識誤り推定ファイルは、信頼度尺度ファイル (CM-file) と非常に類似したフォーマットとなっている。CM-file との相違点は、各単語ごとの出力において、第 6 カラム以降に信頼度尺度ではなく推定結果を示す値が格納さ

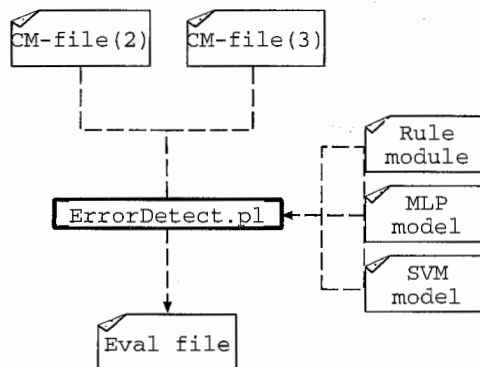


図 A.2: 認識誤り検出ツールデータフロー (検出)


```

:
11:10101:10101:c:C:1:1.42887273447078
11:6:6:c:C:-:-
Utterance:12
12:5:5:c:C:-:-
12::10427:i:E:1:0.70015641148866
12:20305:10501:s:E:0:-0.999051861363199
12:10046:10046:c:-:-:-
12:10203:10203:c:C:1:0.999702282292357
12:10014:10014:c:C:1:1.11065420478367
12:10039:10039:c:C:1:1.91550701501151
12:10021:10021:c:C:1:1.47392707122701
12:6:6:c:C:-:-
Utterance:13
13:5:5:c:C:-:-
13:22911:22911:c:C:1:1.83691369932599
:

```

図 A.4: Eval-file の例

A.3 音声認識誤り訂正フロー

本節では、用例を用いた音声認識誤り訂正ツールについての処理フローを解説する。ここでは、多数の用例ファイルから用例データベースを生成するツール、誤り文に対する類似文検索ツール、誤り箇所代替候補生成ツール、訂正性能評価ツールについて述べる。

A.3.1 用例データベースの構築

誤り訂正で用いる用例データベースを構築する。このためには、用例文が ANSWER ファイルのフォーマットで書かれたものが1つ以上必要である。MkExampleDB.pl は、これらのファイルから用例文を読み込んで、用例文データベースの形にして出力する。MkExampleDB.pl は、内部では、各用例文を正規化ルールによって正規化し、重なり文をマージした後でデータベースに出力する。この時、各用例文は元々の単語 ID(文字列扱い)の系列としてではなく、数値化された単語 ID の系列に変換さる。MkExampleDB.pl の出力ファイルは、したがって、用例文のリストと単語 ID 変換テーブルの2つのファイルとなっている。

この処理のイメージを、図 A.5 に示す。

A.3.2 類似用例文検索

ChkExample.pl は、認識誤りを含んだ認識結果文に対して、認識誤り箇所をマスクした場合の類似用例を用例文データベースから検索する。このコマンドは、あくまでも類似用例文を示すだけで、誤り箇所に対してどの単語を当てはめるべきかの判断は行なわない。次節で述べる MkHypothesis.pl の動作チェック的な意味合いで用いる。

ChkExample.pl は、図 A.6 に示すように、前節で作成した用例データベースを用いて、入力ファイル

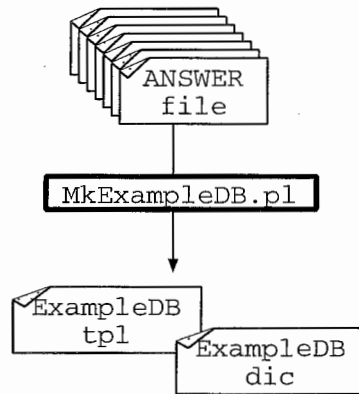


図 A.5: 用例データベース構築

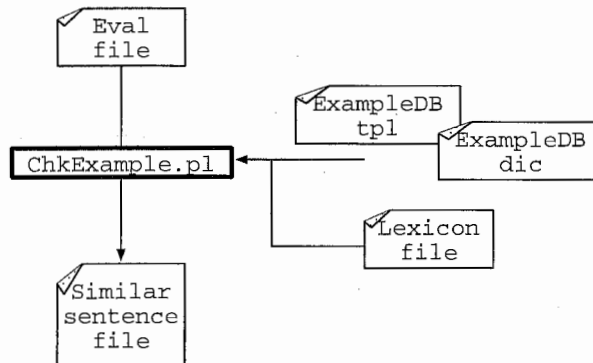


図 A.6: 類似用例文検索

に与えられた認識結果文の類似文を検索する。入力ファイルのフォーマットは、A.2節で述べたコマンド **ErrorDetect.pl** が出力するものと同じフォーマットで、正解単語、認識結果単語、正誤フラグが記されたものとなっている。

A.3.3 認識誤り訂正候補の生成

音声認識の誤り箇所に対する訂正候補を生成するため、**MkHypothesis.pl** を用いる。このコマンドは、**ChkExample.pl** と同様に誤り箇所をマスクした認識結果文に対して、用例データベースから類似用例を検索し、これを元に誤り箇所の代替候補を生成する。

図 A.7 に示すように、**MkHypothesis.pl** は **ChkExample.pl** と同じファイルを入力ファイルとして受け取るが、出力ファイルは、認識誤り箇所に対する訂正候補が、スコア順にリストされたものとなっている。

A.3.4 誤り訂正の集計

MkHypothesis.pl の出力は、各認識誤り箇所に対して代替候補のリストを生成するだけであり、誤り訂正性能の直接的な評価は行なえない。**MkHypothesis.pl** の出力をカウントして、どの程度の数の代替候補が生成されたか、訂正候補中に正解候補が含まれる認識誤りは何か所であったか、訂正候補中から正しく正解候補を選んだ誤り箇所は何か所であったか、などをカウントするため、**CorrectionEval.pl** を用いる。

A.3.5 生成ファイルのフォーマット

similar sentence file

認識誤りを含んだ認識結果文に対して、その類似用例とみなされる用例文が、認識結果文との単語の対応関係と共に出力されるファイルが“similar senece file”である。このファイルは **ChkExample.pl** によって生成される。この出力例を図 A.8 に示す。

ここに示されるように、出力は、各発話の第1行目に認識結果文が示される。この時、認識誤り箇所は <Error> でマスクされており、その誤認識部分に対応する正解単語が、第2行目に単語 ID: 単語表記のスタイルで出力される。第3行目は、類似用例の検索において窓掛けを行なった場合の窓内の単語列を示しており、窓掛けを行っていない場合には第1行目と同じ単語系列となる。第4行目以降が用例集から検索された類似用例文で、出力順は、認識結果文との文距離で昇順にソートされた順番となっている。これら行の各カラムの意味は、以下ようになる。まず第1カラムは、認識結果文と用例文の DP 距離を意味する。第2カラムの文字列は、認識結果文と用例文の DP マッチングによる対応関係を示しており、

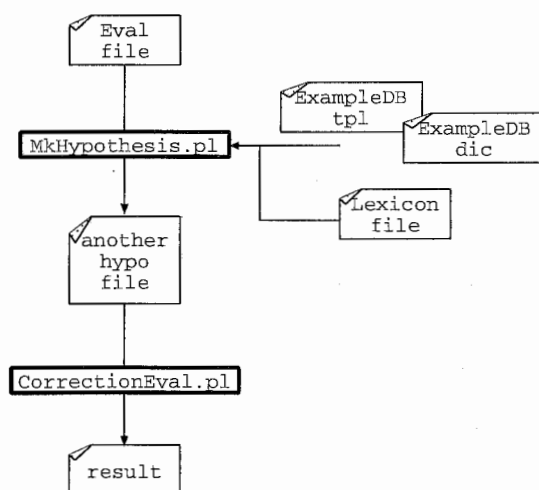


図 A.7: 認識誤り訂正候補の生成

認識結果文を基準として“C”は同一単語を、“S”は置換誤り、“I”は挿入誤り、“D”は脱落誤りを意味している。第3カラム以降は、用例文の単語列がスペースで区切られて出力されている。

another hypo file

“another hypo file”には、認識結果の誤り箇所に対して、用例コーパスの類似文から抽出した候補単語がその文類似度、音韻類似度と共に出力される。このファイルは、MkHypothesis.plによって生成される。この出力例を図A.9に示す。

この例に示す通り、各発話ごとの出力のうち、第1行目～第3行目までは前節で述べた類似用例文出力(similar sentence file)の場合と同じ意味となる。第4行目以降が類似用例文から抽出した候補単語の出力となっており、その出力順は、文距離と音韻的距離の重み付き和によって得られたスコアの順番となっている。また、同一単語が2回以上抽出された場合には、スコアの最適なもの1つにマージしている。各候補単語の出力行での、“:”で区切られた各カラムの意味は、次のようになっている。まず第1カラムは、当該単語を抽出する元となった類似用例文と、認識結果文との文距離である。第2カラムは、候補単語の音韻系列と認識誤り区間の音韻系列の音韻距離である。繰り返しになるが、各候補単語は第1カラムと第2カラムの値の重み付き和によってソートされている。第3カラムは候補単語の単語IDが第4カラムには単語表記がそれぞれ出力される。また上記例にはないが、誤り区間が複数の単語から成るときには、それらの単語を“,”で連結したものが出力される。

```
[6] どの 便 <Error> 乗れますか
10205: なら
どの 便 <Error> 乗れますか
0.17 (CCSCCC): どの 便 に 乗れますか
0.17 (CCSCCC): どの 便 なら 乗れますか
0.50 (CSSSCC): どの カード が つかえますか
0.50 (ISSCCC): サボイホテル から 乗れますか
0.50 (SSSCCC): この ホテル から 乗れますか
0.50 (CSSSCC): どの スタイル に なさいますか
0.50 (IISCCC): みんな 乗れますか
```

図 A.8: similar sentence file の例

```
[6] どの 便 <Error> 乗れますか
10205: なら
どの 便 <Error> 乗れますか
0.167:1.00:10051: に
0.167:1.50:10205: なら
0.500:0.00:10104: が
0.500:0.50:10103: は
0.500:1.00:10067: で
0.500:1.00::
0.500:1.00:10017: を
0.500:1.50:10377: 人
```

図 A.9: another hypo file の例

A.4 各ツールについて

本節では、これまで述べてきた認識誤り検出、認識誤り訂正に関連する以下のツール群について、そのコマンドラインオプション、機能などを詳述する。

- 誤り検出関連
 - CalcCM.pl
 - CMMerge.pl
 - ReFlag.pl
 - ExtMeasure.pl
 - ErrorDetect.pl
 - LinearDisc.pl
 - LinearDisc2.pl
 - MkDecisionRule.pl
 - MkMLPProj.pl
- 誤り訂正関連
 - MkExampleDB.pl
 - ChkExample.pl
 - MkHypothesis.pl
 - CorrectionEval.pl

NAME CalcCM.pl – 認識結果の各単語について信頼度尺度を計算し、出力する

SYNOPSIS

```
CalcCM.pl [-L lexicon] [-u utterance#] [-l lattice-ext.] [-r result-ext]
          [-MO module-option] [-dv] [-dh] basename
CalcCM.pl [-L lexicon] [-u utterance#] [-MO module-option] [-dv] [-dh] lattice,result
```

DESCRIPTION

CalcCM.plは、ATRallやATRresultの出力する *lattice* ファイルおよび *result* ファイルを入力ファイルとして、認識結果の各単語ごとの信頼度尺度を“CM-fileフォーマット”で標準出力に出力する。ただし、ここで言う *lattice* ファイルや *result* ファイルは、オリジナルのATR-SPRECが生成するファイルを指すのではなく、改造版ATRallやATRresultによって出力される、情報が追加されたファイルを指す。

本ツールを利用するには、1組の *lattice* ファイルと *result* ファイルを指定しなければならない。その指定方法は2通りある。第1の方法は、*lattice* ファイルと *result* ファイルの拡張子(通常は .lat と .res)を削った *basename* のみを与える方法、もう1つは、2つのファイルの名前を,(カンマ)で連結して与える方法である。

本ツールにおいて、実際の信頼度尺度計算部分は、perlモジュールの形式でモジュール化されている。このモジュールは、信頼度尺度のカテゴリごとに複数用意されており、いずれのモジュールをインポートして走らせるかによって、本ツールの出力内容が変化する。利用する信頼度尺度計算モジュールの指定方法は、現状では、CalcCM.plスクリプトの冒頭部分で多数羅列されているモジュールの中から1つだけ選択して、そのコメントを外す(行頭の#を削除する)ことである。複数の信頼度尺度計算モジュールをインポートしてしまった場合、基本的には、最後にインポートされたモジュールの機能が有効になるが、モジュールの競合により予期せぬ動作をする恐れもある。モジュールのインポートのためにCalcCM.plを編集する際には、注意が必要である。

本ツールは、以下のようなコマンドラインオプションを取る。

- L 認識辞書 (*lexicon*) ファイルの指定を行なう。本ツールを利用する上で必ず必要なファイルである。デフォルト値は“./ldata/LEX2.W”となっている。
- u 発話番号の指定を行なう。このオプションで指定がなかった場合、*lattice* ファイルと *result* ファイル中で得られている、全ての認識結果について信頼度尺度の計算を行なうが、特定の発話についてのみ結果が欲しい時に本オプションを利用する。指定発話が複数ある場合には、それらの番号を,(カンマ)で連結した文字列を与えればよい。
- l *basename* から *lattice* ファイルの名前を合成する際に、デフォルトでは lat が *lattice* ファイルの拡張子であると仮定している。これを異なる拡張子に変更したい場合に、本オプションを指定する。
- r *basename* から *result* ファイルの名前を合成する際に、デフォルトでは res が *result* ファイルの拡張子であると仮定している。これを異なる拡張子に変更したい場合に、本オプションを指定する。
- MO 信頼度尺度計算のためにインポートする、モジュールに依存したオプションを指定する。オプションの種類や目的は、インポートするモジュールに依存しているため、これについては別途記述する。
- dh デバッグ目的で、内部情報を出力する。
- dv 連結語(+による連結語)を個別に分解して、信頼度尺度を計算する。全ての信頼度尺度計算モジュールについて、この機能が正しく動作する訳ではない。単語ベースでの信頼度尺度を取得したいなら、むしろ連結語なしでATR-SPRECを走らせた結果を利用することを強くお勧めする。

CAUTION

本ツールは、信頼度尺度計算モジュールを実行時に指定する手段を持たない。その都度、**CalcCM.pl**を編集する必要がある。

CalcCM.plの冒頭部で、2つ以上の信頼度尺度モジュールを有効にしてしまった場合、その動作は保証されない。

SEE ALSO

A.2.8節, A.5節

NAME CMMerge.pl - 複数の信頼度尺度ファイルをマージして、1つの信頼度尺度ファイルを作る

SYNOPSIS

CMMerge.pl CM-file:#... [CM-file:#... ...]

DESCRIPTION

CMMerge.pl は、複数の信頼度尺度ファイル (CM-file) を読み込んでその中から指定された尺度を選択、まとめて1つの“CM-file”フォーマットのデータとして、標準出力に出力する。このようなツールを用意したのは、CalcCM.pl が1度に1つの信頼度尺度計算モジュールしか利用できないため、異なるカテゴリの尺度 (異なる計算モジュールによる結果) をまとめて、学習機械などで利用することが困難であったためである。

本ツールによってマージされる複数の信頼度尺度ファイル (CM-file) は、全て同一の認識結果ファイル (*lattice, result*) から生成されたものでなければならない。そうでなかった場合、本ツールの出力するデータは、意味をなさない。

本ツールは、最低でも1つの信頼度尺度ファイルと、そこから抽出する信頼度尺度のカラム番号を指定しなければならない。その指定フォーマットは、以下の通り。

```
<指定> ::= CM-file:<抽出カラム>
<抽出カラム> ::= <カラム指定> | <抽出カラム>,<カラム指定>
<カラム指定> ::= <カラム> | <カラム範囲>
<カラム範囲> ::= <カラム>-<カラム>
<カラム> ::= <0以上の整数>
```

たとえば、

```
cm_file1:5,6,10-15
```

は、“cm_file1”なる信頼度尺度ファイルから、第5,6,10,11,12,13,14,15カラムの値を抽出し、これを新たな信頼度尺度ファイルに出力することを意味する。上記のような記述を、複数ファイルについて記述すれば、出力はそれらが1つのファイルにまとめられたものとなる。

1つのファイルの“抽出カラム”についての記述中に、同一カラムが複数回指定されていても、出力は1度だけである。すなわち、“cm_file:5,5-6”と“cm_file:5,6”は同じ出力となる。また、同一ファイルについての記述が複数に分けてコマンドラインで指定されても、所望の動作をする。すなわち、“cm_file:5 cm_file:6”と“cm_file:5,6”は同じ出力となる。

出力の順番は、コマンドラインに指定された順番が保持される。すなわち、“cm_file:5,6”と“cm_file:6,5”は同じ出力ではない。同様に、“cm_file1:5 cm_file2:5”と“cm_file2:5 cm_file1:5”も同じではない。

CAUTION

上述のカラム指定では、perlの配列流のカウントを行なって、0から始まるものとする。すなわち、第1カラムはカラム番号0であり、第nカラムはカラム番号n-1である。

1つの指定中で、同一カラムの抽出指定があるかどうかをチェックする機構は入っているが、複数の指定にまたがった場合には、この機構は機能しない。すなわち、“cm_file:5,5”は“cm_file:5”と同じであるが、“cm_file:5 cm_file:5”は単独の“cm_file:5”と同一とはならない。

SEE ALSO

CcalCM.pl, A.2.8節

NAME ReFlag.pl - 認識誤り検出の評価対象外単語にフラグをつける

SYNOPSIS

```
ReFlag.pl [-M merge] [-c condition] [-n] [-d] [CM-file ...]
```

DESCRIPTION

認識結果の単語のうち、特定の条件にある単語は、認識誤り検出のための学習機械のトレーニングおよび評価に用いないようにしたい場合に、それらの単語に不要フラグを付ける目的で **ReFlag.pl** を利用する。(例えば、間投詞の除外など。)

本ツールは、コマンドラインに指定された1つ以上の *CM-file*、または標準入力から読み込んだ“CM-file”フォーマットのデータを入力として、指定された条件に該当する単語について、不要フラグを付した新たな“CM-file”フォーマットのデータを標準出力に出力する。

本ツールに2つ以上の *CM-file* を指定した場合、これらのファイルを連結したのものを、**ReFlag.pl** の入力としたのと同じ結果になる。

本ツールのコマンドラインオプションは以下の通り。

- M 間投詞マージファイル (*merge*) を指定する。これにより、*merge* ファイル中に記述のあった単語は、除外単語としてトレーニング・評価の対象から除外される。(複数指定可)
- c 本ツールによってフラグを付す、除外単語の条件を記述する。ここに記述された条件に合った単語は、除外単語とみなされることになる。複数の指定も可能である。本オプションで指定する条件は、(ほぼ)そのまま perl によって評価されるので、perl の文法に合った記述であれば、どのような条件を書いてもよい。ただし、*CM-file* 中の各カラムの値を表わす変数として、(awk 風に) \$<数値> が用意されている。例えば “\$5” は、*CM-file* 中の第6カラムの値を意味している。(perl の配列インデックスは0から始まるため、1つずれていることに注意)
一例として “(\$5 > 0) && (\$6 < 0)” は、*CM-file* の第6カラムの値が正で、第7カラムの値が負となっているような単語は、除外単語としてのフラグを付与するよう指定している。
- n デフォルトで本ツールは、信頼度尺度の値が“-”となっているような単語については、無条件で除外単語のフラグを付与する。このような機能を停止し、他の単語と同様に評価させる必要がある場合にこのオプションを与える。
- d デバッグ出力を生成する。

CAUTION

SEE ALSO

A.2.8節

NAME ExtMeasure.pl - CM-file から学習機械トレーニング・評価用のデータを生成する

SYNOPSIS

```
ExtMeasure.pl -c columns ... [-f format] [-o basename] [-C correct-char] [-E error-char]
[-e column] [-b] [-T Titles] [-t] [CM-file ...]
```

DESCRIPTION

ExtMeasure.pl は、ニューラルネットや決定木といった学習機械をトレーニングしたり、トレーニングされた学習機械を評価したりするためのデータを生成する。

本ツールは、コマンドラインに指定された *CM-file*、または標準入力から読み込んだ “CM-file” フォーマットのデータからデータを抽出、整形する。整形フォーマットは、以下の4種類の学習機械に対応している。

線型判別器 自家製の **LinearDisc.pl** で学習。

決定木 **c5.0** で学習・評価。

ニューラルネット ATR製 **Dmlptrain** で学習。**Dmlptest** で評価。

SVM 奈良先端製 **TinySVM** を利用。**svm_learn** で学習。

本ツールのコマンドラインオプションは以下の通り。

-c *CM-file* から抽出するデータの列を指定する。指定の方法は、以下の通り。

```
<指定> ::= <抽出カラム> | <時間指定>:<抽出カラム>
<時間指定> ::= <整数>
<抽出カラム> ::= <カラム指定> | <抽出カラム>,<カラム指定>
<カラム指定> ::= <カラム> | <カラム範囲>
<カラム範囲> ::= <カラム>-<カラム>
<カラム> ::= <0以上の整数>
```

このように、基本的には **CMMerge.pl** で述べたカラム指定方法と同じ記述で、抽出カラムの指定ができる。しかしここでは記述が拡張されており、時間遅延成分の抽出が可能となっている。すなわち、カラム指定の前に時間指定を:(コロン)で連結して与えることができる。たとえば、“-c-1:5 -c5” のような指定を行なった場合、当該単語の第6カラムの信頼度尺度だけでなく、その直前の単語の信頼度尺度も同時に抽出することを意味する。これは、TDNN(Time Delayed Neural Network) のような、前後コンテキストを考慮したモデルを学習・評価するためのデータを生成するためのものである。

このオプションは、複数指定可能である。

-f *format* に与えた数値によって、出力フォーマットを指定する。数値として許されるのは 0, 1, 2, 3 の4種類の値で(デフォルトでは“0”)、それぞれ次の用途向けの出力を行なう。

0 線型判別用フォーマット。標準出力に出力される。**LinearDisc.pl**, **LinearDisc2.pl** で利用可能。

1 決定木 **c5.0** 用のフォーマット。出力ファイルは2つになるため、引数の **-o** オプションにそれらのファイルのベース名 (*basename*) を必ず与える。実際に生成されるファイルの名前は、*basename* に拡張子 **.name** と **.data** を連結したものとなる。

2 ニューラルネット **Dmlptrain**, **Dmlptest** 用のフォーマット。このオプションによる出力も複数のファイルになるため、**-o** オプションによる出力ファイルのベース名 (*basename*) 指定が必須となる。出力ファイル名は、*basename* に拡張子 **.in** と **.out** を連結したものとなる。

3 SVM(**TinySVM(svm_learn)**) 用のフォーマット。出力ファイルは1つであるため、標準出力にそのまま出力される。

- o 前述の出力フォーマットオプション指定 (-f) によっては、出力ファイルが、複数のファイルになる場合がある (-f1 or -f2)。このような場合に、出力ファイルの名前を指定する。実際に生成されるファイルは、引数に指定された basename に拡張子を追加した名前となる。(-f により必須オプション)
- C CM-file 中の単語の正解・誤りを示すフラグ (第4, 第5カラムに記述) のうち、認識正解を意味するフラグ文字を指定する。デフォルトでは “c” が正解文字となっているが、これとは異なる文字または複数の文字を正解フラグとしたい場合に指定する。(例: -CcC ... “c” と “C” を正解フラグと見なす。)
- E CM-file 中の単語の正解・誤りを示すフラグ (第4, 第5カラムに記述) のうち、認識誤りを意味するフラグ文字を指定する。デフォルトでは “sid” が誤認識フラグとなっているが、これとは異なる文字または誤認識フラグを追加したい場合に指定する。(例: -Eside ... “s,i,d” に加えて “e” も誤認識フラグとして追加する。)
- e 認識結果の正解・誤りのフラグが記述されているカラムを指定する。デフォルトでは、第4カラムを意味する “3” となっている。
- b データカラムの値が ”(空)” になっている場合、本ツールは当該単語に関するデータの出力を行なわない。この機能を停止し、常に出力させる場合には、このオプションを指定する。
- T オプション “-f0” による出力の際に、出力の第1行目に各カラムの意味を記述するタイトル行を追加することができる。(<文字列>, <文字列>, ... のスタイルで指定する。) 歴史的経緯により存在するオプション。
- t 入力ファイル CM-file の第1行目にタイトル行がある時、これを無視させるオプション。歴史的経緯により存在するオプション。通常は必要ない。

CAUTION

SEE ALSO

CMMerge.pl, LinearDisc.pl, LinearDisc2.pl, A.2.8節

NAME ErrorDetect.pl - 学習されたモデルを用いて、CM-file中の各単語の正解・誤りを判別する

SYNOPSIS

```
ErrorDetect.pl -c columns ... [-m type] [-M model] [CM-file ...]
```

DESCRIPTION

ErrorDetect.pl は、引数に指定された学習機械モデルファイルまたは、カレントディレクトリにある判別ルール perl モジュールを用いて、CM-file中の各単語の正解・誤りを判別し、その結果を“Eval-file”フォーマットで標準出力に出力する。入力データは、コマンドラインに1つ以上の CM-fileを与えるか、標準入力から“CM-file”フォーマットのデータを与えるものとする。

本ツールがいずれのモデルファイルまたは判別ルールモジュールを利用するかは、コマンドラインオプション“-m”によって決まる。

本ツールが対応している学習機械は、**ExtMeasure.pl**と同様で、以下の4種類である。

- 線型判別器** 自家製の **LinearDisc.pl**により学習。
生成された判別ルールモジュールを利用する。
- 決定木** **c5.0**で学習。出力を **MkDecisionRule.pl**によって整形して作った、判別ルールモジュールを利用する。
- ニューラルネット** ATR製 **Dmlptrain**で学習。このツールが出力するモデルファイルを利用する。
- SVM** 奈良先端製 **TinySVM**を利用。 **svm_learn**で学習。
出力されたモデルファイルを利用する。

本ツールでの入力ファイルとなる CM-fileは、利用するモデルファイルまたは判別ルールファイルを生成するために用いた CM-fileと、データの並びが全く同じものでなければ意味をなさない。そうでなかった場合、本ツールの出力は正しくない結果となる。

本ツールのコマンドラインオプションは以下の通り。

- c CM-fileから抽出するデータの列を指定する。指定の方法は、**ExtMeasure.pl**と全く同じで、以下の通り。

```
<指定> ::= <抽出カラム> | <時間指定>:<抽出カラム>
<時間指定> ::= <整数>
<抽出カラム> ::= <カラム指定> | <抽出カラム>,<カラム指定>
<カラム指定> ::= <カラム> | <カラム範囲>
<カラム範囲> ::= <カラム>-<カラム>
<カラム> ::= <0以上の整数>
```

本オプションにより抽出されるデータの並びは、モデルファイル・判別ルールファイルを生成するのに利用したデータと、同じ並びであらねばならない。そうでなかった場合、本ツールの出力は正しい出力とはならない。(複数指定可)

- m typeに与えた数値によって、判別に利用するモデルのタイプを指定する。数値として許されるのは0, 1, 2, 3の4種類の値で(デフォルトでは“0”)、それぞれ次のモデルを利用する。

- 0 線型判別による判別ルールモジュールを利用(デフォルト)。このオプションを指定した場合、カレントディレクトリに **DecisionRule.pm**という名前の判別ルールモジュールが存在しなければならない。このファイルは、**LinearDisc.pl**に“-02”オプションを与えることで得ることができる。
- 1 決定木 **c5.0**により生成した判別ルールモジュールを利用。“-m0”の場合と同様、この場合にもカレントディレクトリに **DecisionRule.pm**という名前の判別ルールモジュールが存在しなければならない。このファイルは、**c5.0**に“-r”オプションを与えて得られた出力を、**MkDecisionRule.pl**で整形することによって得られる。

2 ニューラルネット学習ツール、**Dmlptrain**によって生成されたモデルファイルを利用。このオプションを指定した場合には、以下に述べる“-M”オプションによって、モデルファイルのパス名を与えてやる必要がある。

3 SVM学習ツール **svm_learn**によって得られたモデルファイルを利用。このオプションを指定した場合には、以下に述べる“-M”オプションによって、モデルファイルのパス名を与えてやる必要がある。

-M オプション“-m”によって、2または3(ニューラルネットまたはSVM)を指定した場合に、そのモデルファイルの名前 model を指定する。

CAUTION

SEE ALSO

ExtMeasure.pl, LinearDisc.pl, MkDecisionRule.pl, A.2.8節, A.2.8節

NAME LinearDisc.pl - 判別分析を行ない、判別式または判別ルールモジュールを生成する

SYNOPSIS

```
LinearDisc.pl [-O type] [-T test-set] [-v] [-t] [data-file ...]
```

DESCRIPTION

LinearDisc.pl は、“**ExtMeasure.pl -f0**”の出力を入力として判別分析を行ない、その結果を出力する。出力としては、入力データに対する統計値(平均・共分散行列)、判別率、判別式(の係数)、判別ルールを perl モジュールフォーマットにしたもの、などである。

本ツールは、“**ExtMeasure.pl -f0**”の出力をそのまま標準入力から読み込むか、出力を格納したファイル(*data-file*)をコマンドラインに指定するかによって動作する。本ツールが出力する情報は、コマンドラインオプションの指定によって異なるが、いずれの出力も標準出力に出力される。

本コマンドで有効なオプションは以下の通り。

- O 出力する情報のタイプを *type* の数値で指定する。有効な数値と出力される情報は、以下の通り。
 - 0 判別分析に関連した各種統計値を出力する。具体的には、各項目の平均値、共分散行列、判別係数、および判別分析による認識正解・誤りの適合率(precision)、再現率(recall)などである。本ツールは、この出力がデフォルト出力である。
 - 1 単に判別係数のリストを出力する。この出力は、**LinearDisc2.pl** でそのままコマンドラインオプションで与えることのできるフォーマットとなっている。
 - 2 **ErrorDetect.pl** が利用する、判別ルール perl モジュールを生成する。出力は標準出力に出力されるが、**ErrorDetect.pl** にインポートさせるために、この出力が“**DecisionRule.pm**”という名前のファイルに格納される必要がある。
- T *test-set* に、テストセットの格納されたファイルを指定すると、このテストセットに対する、正解・誤り判別を行なう。このテストセットファイルは、“**ExtMeasure.pl -f0**”により生成したデータでなければならない。この機能は、“-00”を指定した時のみ機能する。
- v 判別分析結果出力において、単に適合率・再現率を報告するだけでなく、個々のデータに対する判別結果も報告する。この機能も“-00”を指定している時のみ有効である。
- t 入力データファイルの先頭行にタイトル行があることを想定する。(歴史的経緯により存在するオプション。通常は無視可。)

CAUTION

SEE ALSO

ExtMeasure.pl, ErrorDetect.pl, LinearDisc2.pl

NAME LinearDisc2.pl - 判別係数を振ることによる検出・再現率の変化を調べる

SYNOPSIS

```
LinearDisc2.pl -c coef-list [-s step] [-d division] [-debug] [-t] [data-file ...]
```

DESCRIPTION

判別分析による正解・誤りの適合率と再現率は、通常トレードオフの関係にある。本ツールは、この様子を調べるためのツールである。具体的には、**LinearDisc.pl**により得られた判別式に対して定数項の値を前後に振って、その都度正解・誤りの適合率と再現率を調べ、これを出力する。

LinearDisc2.plは、引数に指定された判別式(判別係数のリスト)を用いて、入力データを判別分析する。入力データは、“**ExtMeasure.pl -f0**”の出力そのものを標準入力から取り込むか、コマンドラインに指定された data-file に、“**ExtMeasure.pl -f0**”出力が格納されているものとして読み込む。

本ツールは、1つの定数項に対する適合率・再現率などをスペースで区切ったリストを1行に1つずつ生成し、これを標準出力に出力する。出力は、一例として次のようになる。

```
      :  
-15.821949 98.97 100.00 100.00 75.00 99.00 75.00  
-14.821949 98.96 98.96 75.00 75.00 98.00 50.00  
-13.821949 98.96 98.96 75.00 75.00 98.00 50.00  
-12.821949 98.95 97.92 60.00 75.00 97.00 25.00  
      :
```

この各カラムの意味は次の通り。

定数項 正の適合率 正の再現率 負の適合率 負の再現率 判別率 誤判別改善率

本コマンドで有効なオプションは以下の通り。

- c 判別式を判別係数のリストの形式で指定する。これは、各係数値を ,(カンマ)で連結したものである。このリストは、“**LinearDisc.pl -01**”の出力をそのまま利用することができる。(必須オプション)
- s 定数項の振り幅を指定する。デフォルトでは 0.1。
- d 結果の出力個数を指定する。デフォルトでは 40。(ただし計算不能な場合は、これよりも少ない個数のデータが出力される。)
- debug 出力において、より詳しい情報を出力する。
- t 入力データファイルの先頭行にタイトル行があることを想定する。(歴史的経緯により存在するオプション。通常は無視可。)

本ツールにおいて定数項は、coef-listに指定された定数項の値を中心値として、前後に division/2 個づつ、step刻みで並べた値を用いるものとする。

CAUTION

数学的にあまり正確な検証はなされていない。

SEE ALSO

LinearDisc.pl, ExtMeasure.pl

NAME MkdirDecisionRule.pl - c5.0による決定木から、判別ルールモジュールを生成する

SYNOPSIS

```
MkdirDecisionRule.pl [-m] [output-file]
```

DESCRIPTION

MkdirDecisionRule.plは、決定木生成ツール **c5.0**の生の出力を、perlの関数として利用可能なように整形する。

このツールは、**c5.0**の出力を標準入力から取り込むか、コマンドラインに指定された output-file に出力が格納されているものとして、perl関数、または判別ルール perlモジュールを生成する。本ツールの出力は、標準出力に出力される。

本ツールが想定している入力データは、**c5.0**に“-r”オプションを与えて得られた出力でなければならない。

本ツールで利用可能なオプションは以下の通り。

- m デフォルトでは、本ツールは **c5.0**の出力から perl関数を生成する。**ErrorDetect.pl**で利用可能なように、perlのモジュールを生成する時本オプションを指定する。
本オプション指定による出力を、**ErrorDetect.pl**で利用するには、出力を“**DecisionRule.pm**”という名前のファイルに格納する必要がある。

CAUTION

SEE ALSO

ErrorDetect.pl

NAME MkMLPProj.pl - MLP ツール (Dmlptrain, Dmlptest) で利用可能なプロジェクトファイルを生成する

SYNOPSIS

MkMLPProj.pl project-name

DESCRIPTION

MkMLPProj.pl は、ニューラルネット学習・評価ツールである Dmlptrain, Dmlptest が利用する、プロジェクトファイルのテンプレートを生成する。

本来このファイル群は手作業で作成するものであるが、“ExtMeasure.pl -f2” の出力するファイルセットから、これを用いて学習・評価するための最低限のプロジェクトファイルを自動生成することが可能である。本ツールは、これを行なうためのものである。

本ツールは、“ExtMeasure.pl -f2” の出力するファイルが project-name.in, project-name.out という名前で作成されていることを前提として、これらのファイルを読み込み、project-name.mlp_def, project-name.list_in, project-name.list_out の3つのファイルを自動生成する。

CAUTION

project-name.mlp_def に記述される内容は、Dmlptrain が最低限動作するための記述がなされているだけである。適宜変更が必要である。

SEE ALSO

ExtMeasure.pl

NAME MkExampleDB.pl - 用例文コーパスから、用例文検索用のデータベースを構築する

SYNOPSIS

MkExampleDB.pl [-L *lexicon*] [-m *merge*] [-u] [-dc] *list-file* *basename*

DESCRIPTION

MkExampleDB.pl は、“ANSWER ファイル”フォーマットで記述された用例文コーパスをマージ・整形して、誤り訂正ツールで利用可能なデータベースを作成する。

本ツールは、最低限次の 2 つの指定がされている必要がある。まず第 1 に、用例文ファイル名をリスト (1 行に 1 ファイル名) とするリストファイル *list-file*。もう 1 つは、生成したデータベースを格納するファイルのベース名 *basename* で、この場合、実際に出力されるファイルの名前は *basename.tpl* と *basename.dic* となる。

本ツールは、以下のようなコマンドラインオプションを取る。

- L 認識辞書 *lexicon* の名前。これは、用例文中の未知語の判別などに用いられる。デフォルトでは、これは “ldata/LEX.TRA+PB.comp” である。
- m 用例文のマージにおいて、ヒューリスティックな正規化ルールによって用例文を正規化するルールを指定する。利用可能なレベルと、その意味は以下の通り。
 - 0 文の正規化処理は一切行わず、単に同一文のみマージする。(デフォルト動作)
 - 1 問投詞の削除による文正規化の後に、同一文のマージをする。
 - 2 “-m1” の正規化に加えて、品詞名が “数詞” であるものを全て同一単語、“数字” であるものを全て同一単語として正規化した後に、同一文をマージする。
 - 3 “-m2” の正規化に加えて、品詞名が “日姓, 日名, 外姓, 外名” であるものを全て同一単語として正規化した後に、同一文をマージする。
 - 4 “-m3” の正規化に加えて、品詞名が “地名” であるものを全て同一単語、“会社名” であるものを全て同一単語として正規化した後に、同一文をマージする。
 - 5 “-m4” の正規化に加えて、品詞名が “日時” であるものを (先行単語が “数詞” である場合には、これも合わせて) 全て同一の単語として正規化した後に、同一文をマージする。
 - 6 “-m5” の正規化に加えて、当該単語が金額単位 “円, ドル, セント, ポンド, ペニー” であり、かつそれに先行する単語の品詞が “数詞” であるものを、まとめて 1 つの同一単語であるとして正規化した後に、同一文をマージする。
 - 7 “-m6” の正規化に加えて、品詞名が “終助詞” であるものは全て同一単語として正規化した後に、同一文をマージする。
- u 用例文コーパス中に、未知語があった場合の対処法を指定する。デフォルト (0) では、未知語のある用例文は無視してデータベースに登録しない。非 0 を指定すると、未知語のある用例文も未知語のまま用例文に登録する。
- dc デバッグ出力として、データベースに登録する (正規化済み) 用例文を、標準出力に出力する。

CAUTION

これまでの実験・評価においては、常にマージレベルを “-m1” としたものを用いてきた。

他のマージレベルで生成したデータベースを用いて、実験・評価を行なった場合に、誤り訂正関連ツールが正常動作するかどうかのチェックは不十分である。

SEE ALSO

NAME ChkExample.pl – 認識結果から、類似用例文を検索、出力する

SYNOPSIS

```
ChkExample.pl -B basename [-L lexicon] [-S sem-file] [-w res-column] [-a ans-column]
[-f flag-column] [-p phone-column] [-W window] [-e error-char] [-n hypo-num]
[-m merge] [-ts threshold] [-te err-count] [-u utterance] [-o format]
[-O misc-opt] [CM-file | Eval-file]
```

DESCRIPTION

ChkExample.pl は、CM-file や Eval-file のような認識結果とその正誤フラグが記述されたファイルを入力として、正誤フラグによって認識誤りがあると判定された認識結果について、その認識結果文と類似する用例文を用例文データベースより検索し、これを文距離の近いものから順に出力する。

このツールは、CM-file または Eval-file (いずれでも可) をコマンドラインに指定するか、これらのファイルと同じデータ形式のデータを標準入力から読み込む。また用例文検索の結果を、標準出力に “similar sentece file” フォーマットで出力する。

本ツールは、用例検索ライブラリ (**TmpMatch**) をインポートして動作する。さらにこの用例検索ライブラリは、実行ディレクトリに “word_dist.so” なる shared object が存在することを想定している。この shared object は、単語間の距離を定義する C 関数である。いくつかの “word_dist.so” のサンプルが C ソース付きで “work/WordDist” の下に格納されている。参考にされたし。

本ツールの目的は、どのような類似用例文を用いて誤り箇所に対する訂正候補の抽出を行なったかを調べるためのものである。(訂正候補の抽出を行なうツールは **MkHypothesis.pl** である。) したがって本ツールは、**MkHypothesis.pl** と全く同じコマンドラインの記述で、全く同じ条件の用例検索を行なう。そのため、本ツールでは全く意味をなさないダミーのオプションがいくつか存在するが、これは **MkHypothesis.pl** とコマンドラインの互換性を保つために用意されている。このようなダミーのオプションを含めて、本ツールで有効なコマンドラインオプションには次のようなものがある。

- B 類似用例文検索のための、用例文データベースファイルセットのベース名 basename を指定する。データベースとして実際に読み込まれるファイルの名前は、“basename.tpl” と “basename.dic” の 2 つのファイルとなる。(必須オプション)
- L 認識辞書 (lexicon) ファイルの指定を行なう。デフォルトでは、“./ldata/LEX.TRA+PB.comp” となっている。
- S 単語意味コードテーブルファイル (sem-file) を指定する。類似用例文検索において、単語意味距離を考慮した用例文検索を行なう時、その意味コードテーブルファイルを指定する。デフォルトでは、“(空)” となっている。これは、単語意味は全く考慮しないことを意味する。
- w 認識結果の単語 ID が記述されているデータカラムを指定する。CM-file, Eval-file のフォーマットに従うならこの値は “2” (第 3 カラム) で、これがデフォルト値である。
- a 認識結果に対する正解単語の単語 ID が記述されているデータカラムを指定する。CM-file, Eval-file のフォーマットに従うならこの値は “1” (第 2 カラム) で、これがデフォルト値である。
- f 単語の正解・誤りフラグが記述されているデータカラムを指定する。もし、正解・誤りフラグとして正解ラベルに基づいたフラグを利用するなら、CM-file, Eval-file ではこの値は “3” (第 4 カラム) となり、これがデフォルト値である。
- p 当該区間の音韻系列が記述されているデータカラムを指定する。CM-file, Eval-file のフォーマットでは、このカラム位置は規定されていない。**CalcCM.pl** 等によって生成した音韻系列記述のあるカラムを指定する必要がある。本オプションは互

換性のためのオプションであり、実際には用例検索において何らの働きもしていないが、データの正当性チェックには用いられているため、正しく音韻系列が記述されているカラムを指定してやる必要がある。デフォルトではこの値は“5”(第6カラム)となっている。

- e 単語の正解・誤りフラグのうち、誤りを意味するフラグ文字を指定する。デフォルトでは“sid”となっており、これは“s, i, d”の3つの文字が誤認識を意味するフラグであると指定するものである。
- W 類似用例検索において、認識結果文全体ではなく、認識誤り個所周辺の系列のみを用いて検索を行なう場合に、その窓幅を設定する。設定方法は“<数値>, <数値>”で、それぞれ認識誤り区間の先行方向の窓幅、認識誤り区間の後続方向の窓幅を意味する。ここで、窓幅 n の窓とは内容語を n 個含む幅の窓であるとする。このオプション指定を行なわなかった場合には、認識結果文全体で用例検索するものとする。
- n 検索された類似用例の中から、文距離の近いものから順に幾つの用例文を出力するかを指定する。ただし出力される類似用例文は、きっかりここで指定した数 hypo-num 個になるとは限らない。これは、同一の文距離を有する類似用例が多数存在することが多いため、実際には、hypo-num 以上で、できるだけ hypo-num に近い用例文数となるように出力される。デフォルトではこの値は“5”である。
- m 用例文および認識結果の正規化ルール(マージレベル)を指定する。この値 merge は、用例文データベース構築時(cf. MkExampleDB.pl)に指定したマージレベルと同一の値でなければならない。そうでなかった場合、類似用例の検索は不当なものとなる。デフォルトではこの値は“0”である。
- ts 類似文として認める、文間距離の上限値を指定する。文間距離がこの値より大きくなる用例文については、たとえそれが上位 hypo-num 個の中に入っているも出力されることはない。デフォルト値は“0.99”である。
- te 認識結果文中の誤り単語の数が err-count 個以上のものについては、そもそも用例検索自体を行わない。この値を設定する。デフォルトでは“99”(実質的に、誤りがあれば常に用例検索を行なう)である。
- u 特定発話についてのみ(かつ、その発話が認識誤りしている時のみ)この発話について用例検索を行なう。複数指定可能で、その場合には発話番号を“,”(カンマ)で連結して指定する。
- o 出力フォーマットを指定するためのものであるが、本ツールではダミーオプション。
- O 追加の詳細オプションを指定する。本ツールで有効なものとしては“-ODPConf”のみである。これは、用例文と認識結果文のDPマッチングのパラメータを設定する。指定方法は“-ODPConf=<1>,<2>,<3>,<4>,<5>”である。各<1>~<5>には数値が入り、次のような意味を持つ。
 - <1> 文DPにおける単語置換ペナルティ。
 - <2> 認識結果文に対する、用例文の単語挿入ペナルティ。
 - <3> 認識結果文に対する、用例文の単語脱落ペナルティ。
 - <4> 端点フリーDP条件。
 - 0 両端点固定のDPを行なう。
 - 1 用例文の端点をフリーとする。すなわち、用例文の文頭・文末の単語脱落コストは0とみなす。
 - 2 認識結果文の端点をフリーとする。すなわち、認識結果文の文頭・文末の単語脱落コストは0とみなす。
 - 3 用例文・認識結果文の両方で、端点フリーとしてDPを行なう。
 - <5> 文間距離正規化ルール。
 - 0 正規化なし。置換・挿入・脱落ペナルティの総和が文間距離。

- 1 用例文の単語長で正規化する。
- 2 認識結果文の単語長で正規化する。
- 3 用例文・認識結果文の単語長の和で正規化する。

なお、本オプションのデフォルト値は“-ODPConf=1,1,1,0,0”である。

CAUTION

オプションのデフォルト値に矛盾がある。すなわち“-ODPConf=1,1,1,0,0”である場合、認識結果文と用例文に一箇所でも相異点があると、文間距離は1以上となる。ところが“-ts0.99”であるため、1以上の文間距離となる用例文は無視される。結果として類似用例は何も出力されないことになる。何らかの類似用例を出力させるためには、いずれかのオプションを変更する必要がある。

SEE ALSO

MkExampleDB.pl, MkHypothesis.pl

NAME Mkhypothesis.pl - 認識誤り個所に対する訂正候補を生成する

SYNOPSIS

```
Mkhypothesis.pl -B basename [-L lexicon] [-S sem-file] [-w res-column] [-a ans-column]  
[-f flag-column] [-p phone-column] [-W window] [-e error-char] [-n hypo-num]  
[-m merge] [-ts threshold] [-te err-count] [-u utterance] [-o format]  
[-O misc-opt] [CM-file | Eval-file]
```

DESCRIPTION

Mkhypothesis.pl は、*CM-file* や *Eval-file* のような認識結果とその正誤フラグが記述されたファイルを入力として、正誤フラグによって認識誤りがあると判定された認識結果について、その認識結果文と類似する用例文を用いて、認識誤り個所に対する訂正候補を生成する。

このツールは、*CM-file* または *Eval-file* (いずれでも可) をコマンドラインに指定するか、これらのファイルと同じデータ形式のデータを標準入力から読み込む。また認識誤り個所に対する訂正候補群を、標準出力に “another hypo file” フォーマットで出力する。

本ツールは、用例検索ライブラリ (**TmpMatch**) をインポートして動作する。さらにこの用例検索ライブラリは、実行ディレクトリに “word_dist.so” なる shared object が存在することを想定している。この shared object は、単語間の距離を定義する C 関数である。いくつかの “word_dist.so” のサンプルが C ソース付きで “work/WordDist” の下に格納されている。参考にされたし。

本ツールで出力する訂正候補群が含まれていた、類似用例文そのものを調べる目的のためには、**ChkExample.pl** が利用可能である。本ツールと、**ChkExample.pl** のコマンドラインの記述法は全く同じで、同一のコマンドラインオプションでは同一の条件での用例検索が行なわれる。したがって、コマンドラインオプションについての解説は、**ChkExample.pl** を参照されたし。以下では、**ChkExample.pl** の記述との相異点・補足説明のみ行なう。

- p 誤り区間の音韻と、訂正候補の音韻の間の音韻距離を計算するための、認識結果音韻系列が記述されたデータカラムを指定する。(ChkExample.plでの“-p”についての解説も参照)
- o 歴史的経緯によるオプション。現在はダミー。
- O 追加の詳細オプションを指定する。本ツールで有効なものとして、次のものがある。
 - ODPConf **ChkExample.pl** に述べたものと全く同じ
 - OWeight 訂正候補スコア付けのための、単語間音韻距離と文距離の重みを指定する。指定方法は “-OWeight=<音韻重み>,<文重み>” である。デフォルトでは、“-OWeight=0.1,0.9” となっている。

CAUTION

オプションのデフォルト値に矛盾がある。詳細は **ChkExample.pl** を参照。

SEE ALSO

MkExampleDB.pl, ChkExample.pl, CorrectionEval.pl

NAME CorrectionEval.pl - 誤り訂正の性能評価を行なう

SYNOPSIS

CorrectionEval.pl [-o *format*] [-L *lexicon*] [-S *sem-file*] [*another-hypo*]

DESCRIPTION

MkHypothesis.pl は、誤り個所に対する訂正候補を抽出し、これをスコア順にソートして出力するが、そらの訂正候補が正解であったかどうかの判定は行なわない。CorrectionEval.pl は、これを行なって、誤り個所の訂正性能を評価する。

本ツールは、MkHypothesis.pl の出力を格納したファイル *another-hypo* をコマンドラインに指定するか、“another hypo file”フォーマットのデータを標準入力から読み込むかして動作する。評価結果は、標準出力に出力される。

本コマンドで有効なオプションは以下の通り。

- o 出力する情報のタイプを *format* の数値で指定する。有効な数値と出力される情報は、以下の通り。

0 実質的に何もしない。入力と同じものが出力される。

1 この指定による出力の例を以下に示す。

```
[201] サラダ <Error> <Error> てい ません
10104,13091: が, き:74:2:9
```

ここに示すように、まず第1行目に認識誤りのあった認識結果文が示され、第2行目以降に認識誤り区間ごとの正解単語(系列)と、訂正候補の正解順位等が示される。すなわち“:”(コロン)をデリミタとして、第1カラムに正解単語の単語ID(列)、第2カラムにその表記(列)、第3カラムには全訂正候補の数、第4カラムには正解候補の順位、第5カラムには正解候補以上のスコアを得た訂正候補の数(同スコア候補のカウント)が示される。

2 この指定による出力の例を以下に示す。

```
232 231 22094
0: 132
1: 68
2: 20
3: 6
4: 3
5: 2
```

この出力形式では、各誤り区間の正解訂正候補の順位の集計値が示される。第1行目は全体の数を示しており、左から全誤り区間の数、1つ以上の訂正候補の得られた誤り区間の数、全訂正候補の総数である。第2行目以降が、各順位の正解候補が幾つ得られたかを示している。例えば上記例では、1位の正解候補が得られた誤り個所が全部で68箇所あったことを示す。なお0位というのは、正解候補が得られなかったことを意味し、上記例では、132箇所の誤り箇所では正解候補が得られなかったことを意味する。

3 この出力例を以下に示す。

```
232 231 22094 99 68
```

この出力形式では、さらにまとめた集計結果が示される。各数値の意味は、左から順に、全誤り区間の数、1つ以上の訂正候補が得られた誤り区間の数、全訂正候補数、訂正候補中に正解候補が含まれていた誤り区間の数、1位の正解候補が得られた誤り区間の数、である。

+4 上記1~3に4を足した値を指定すると、訂正候補の正解判定を単語IDで行なうのではなく、品詞IDで行なうようになる。これは、品詞レベルでどの

程度正解を予測できているかを調べる目的のものである。これらの値を指定した時、以下に挙げる“-L”は必須オプションとなる。

- L lexiconに、認識辞書ファイル名を指定する。上記“-o”オプションにおいて、4以上の値を指定した時の必須オプションである。“-o”において3以下の値を指定している時には、このオプション指定は、時間の無駄である。
- S 意味コードファイル sem-file 中では、いくつかの単語品詞の正規化を行なっている。他との互換性のため、この正規化も行ない時には“-L”に加えてこのオプションも指定する。

CAUTION

SEE ALSO

MkHypothesis.pl

A.5 信頼度尺度計算モジュールについて

前述したように、**CalcCM.pl**において、実際に信頼度尺度の計算を行なっている部分は、モジュール化されている。**CalcCM.pl**の出力は、インポートした信頼度尺度計算モジュールに依存する。ここでは、この信頼度尺度計算モジュールについて解説する。

CalcCM.plがインポートする全ての信頼度尺度モジュールは、“ErrorDetection/bin/ConfMeasure”の下に置かれている。

A.5.1 CMNull

何もしない。新たな信頼度尺度計算モジュールを作成する時のテンプレートとして用いる。

A.5.2 CMname

認識結果の各単語の単語表記と音素表記を取り出す。**CalcCM.pl**出力の第6カラム以降の意味は次のようになる。

6カラム 単語表記
7カラム 音素表記

A.5.3 CMPOS

認識結果の各単語の品詞情報を取り出す。**CalcCM.pl**出力の第6カラム以降の意味は次のようになる。

6カラム 単語正規化表記
7カラム 単語品詞表記
8カラム 単語品詞ID

A.5.4 CMTagging

認識結果の各単語のMASTER_LEXICONでのタギング情報を出力する。**CalcCM.pl**出力の第6カラム以降の意味は次のようになる。

6カラム 単語正規化表記
7カラム 単語品詞表記

A.5.5 CMClassID

認識結果の各単語の言語モデル中でのTo-Class IDおよびFrom-Class IDを取得する。**CalcCM.pl**出力の第6カラム以降の意味は次のようになる。

6カラム 当該単語のFromクラス
7カラム 当該単語のToクラス
8カラム 当該単語の先行単語のFromクラス

A.5.6 CMWordLeng

認識結果の各単語の音素長およびモーラ長を計算する。**CalcCM.pl**出力の第6カラム以降の意味は次のようになる。

6カラム 当該単語の単語音素標記
7カラム 当該単語の音素系列長
8カラム 当該単語のモーラ系列長

A.5.7 CMScore

認識結果の各単語の各種スコアを計算する。**CalcCM.pl**出力の第6カラム以降の意味は次のようになる。

- 6カラム 当該単語の 区間始端時間
- 7カラム 当該単語の 区間終端時間
- 8カラム 当該単語の トータルスコア
- 9カラム 当該単語の 音響スコア
- 10カラム 当該単語の 区間時間平均の音響スコア
- 11カラム 当該単語の 言語スコア
- 12カラム 当該単語の クラス 2gram スコア
- 13カラム 当該単語の 単語出力スコア
- 14カラム 当該単語の 10カラムの値 + 11カラムの値

A.5.8 CMBBackScore

認識結果の各単語の 後ろ向き言語スコアを計算する。この計算のため後ろ向き言語モデルが必要である。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

- 6カラム 当該単語の 後ろ向き 2gram スコア
- 7カラム 当該単語の 後ろ向きクラス 2gram スコア
- 8カラム 当該単語の 後ろ向き単語出力スコア

A.5.9 CMBasic

認識結果の各単語と時間的に競合する候補の、種々の競合数をカウントする。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

- 6カラム 当該単語区間に全区間が含まれる、競合候補ののべ数
- 7カラム 当該単語区間に全区間含まれる、競合単語の種類数 (同一単語マージ)
- 8カラム 当該単語区間に単語始端が含まれる、競合候補ののべ数
- 9カラム 当該単語区間に単語始端が含まれる、競合単語の種類数 (同一単語マージ)
- 10カラム 当該単語区間に単語終端が含まれる、競合候補ののべ数
- 11カラム 当該単語区間に単語終端が含まれる、競合単語の種類数 (同一単語マージ)
- 12カラム 当該単語区間と単語区間が重なる、競合候補ののべ数
- 13カラム 当該単語区間と単語区間が重なる、競合単語の種類数 (同一単語マージ)
- 14カラム 当該単語区間を完全に覆う、競合部分パスの総数
- 15カラム 当該単語区間を完全に覆う、競合部分パスの種類数 (同一単語マージ)

A.5.10 CMDivideMerge

認識結果の各単語と同一時刻の単語始端または終端を有する競合候補の、種々の競合数をカウントする。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

- 6カラム 当該単語アークと終端ノードを同じくするアークの数
- 7カラム 当該単語と同一単語IDで、終端時間を同じくするアークの数
- 8カラム 当該単語と終端時間を同じくする単語の種類数 (同一単語マージ)
- 9カラム 当該単語と終端時間を同じくするアークの数
- 10カラム 当該単語と同一単語IDで、終端時間が±30msの誤差内で一致するアーク数
- 11カラム 当該単語と、終端時間が±30msの誤差内で一致する単語種類数 (同一単語マージ)
- 12カラム 当該単語と、終端時間が±30msの誤差内で一致するアーク数
- 13カラム 当該単語アークと始端ノードを同じくするアークの数
- 14カラム 当該単語と同一単語IDで、始端時間を同じくするアークの数
- 15カラム 当該単語と始端時間を同じくする単語の種類数 (同一単語マージ)
- 16カラム 当該単語と始端時間を同じくするアークの数
- 17カラム 当該単語と同一単語IDで、始端時間が±30msの誤差内で一致するアーク数
- 18カラム 当該単語と、始端時間が±30msの誤差内で一致する単語種類数 (同一単語マージ)
- 19カラム 当該単語と、始端時間が±30msの誤差内で一致するアーク数

A.5.11 CMEndState

上記、CMDivideMergeのサブセット。認識結果単語区間の終端時間に着目して、この時点における競合候補数をカウントする。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

- 6カラム 当該単語アークと終端ノードを同じくするアークの数
- 7カラム 当該単語と同一単語IDで、終端時間を同じくするアークの数
- 8カラム 当該単語と終端時間を同じくする単語の種類数(同一単語マージ)
- 9カラム 当該単語と終端時間を同じくするアークの数
- 10カラム 当該単語と同一単語IDで、終端時間が±30msの誤差内で一致するアーク数
- 11カラム 当該単語と、終端時間が±30msの誤差内で一致する単語種類数(同一単語マージ)
- 12カラム 当該単語と、終端時間が±30msの誤差内で一致するアーク数

A.5.12 CMStripHypo

認識結果の各単語と時間的に競合する候補の、種々の競合数をカウントする。“CMBasic”との相違点は、単語競合の条件として時間的なオーバーラップと共に音素系列の類似性も考慮するという点である。たとえば、「京都」と時間的に競合する候補が「今日+と+明日」であっても、音素系列の類似性から競合候補は「今日+と」とみなした上で、競合数をカウントする。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

- 6カラム 当該単語アークと競合するアーク(部分パス)の数
- 7カラム 当該単語と競合する音素系列の種類数(音素系列マージ)
- 8カラム 当該単語と競合する先行単語を含んだ単語系列の種類数(単語系列マージ)
- 9カラム 当該単語と競合する単語系列の種類数(単語系列マージ)

A.5.13 CMStripHypo3

“CMStripHypo”の出力する情報に、当該候補と競合候補のスコア差の情報を追加したセットを出力する。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

- 6カラム 当該単語アークと競合するアーク(部分パス)の数
- 7カラム 当該単語と競合する音素系列の種類数(音素系列マージ)
- 8カラム 当該単語と競合する先行単語を含んだ単語系列の種類数(単語系列マージ)
- 9カラム 当該単語と競合する単語系列の種類数(単語系列マージ)
- 10カラム 当該単語のトータルスコア(音響+言語)と、これと競合する最もスコアの高い部分単語系列とのスコアの差
- 11カラム 当該単語の音響スコアと、これと競合する最もスコアの高い部分単語系列との音響スコアの差
- 12カラム 当該単語の言語スコアと、これと競合する最もスコアの高い部分単語系列との言語スコアの差
- 13カラム 第10カラムのスコア差を、単語の長さで正規化した値
- 14カラム 第10カラムのスコア差を、単語の長さ(フレーム長)で正規化した値
- 15カラム 第10カラムのスコア差を、単語の長さ(単語長)で正規化した値

A.5.14 CM_Hypo_Dist

当該単語と音韻的に近い競合単語(系列)との音響スコアの差を求める。一部、“CMStripHypo”と同じ情報が出力されている。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

- 6カラム “CMStypHypo”による第6カラムと同じ
- 7カラム “CMStypHypo”による第7カラムと同じ
- 8カラム “CMStypHypo”による第9カラムと同じ
- 9カラム 当該単語との競合候補のうち、音素系列が最も類似する候補との音素編集距離
- 10カラム 当該単語との競合候補のうち、音素系列が最も類似する候補との音響スコアの差
- 11カラム 当該単語との競合候補のうち、スコアの最も高い候補との音素編集距離
- 12カラム 当該単語との競合候補のうち、スコアの最も高い候補との音響スコアの差

A.5.15 CMPosteriori

単語グラフを用いて、当該単語の事後確率を求める。このモジュールを正しく動かすためには、ラティス中の各ノードにおける前向き確率と後向き確率を求めておく必要がある。(A.2.1節, 表 A.1 参照。) このモジュールは、次の7つの尺度を1組とした幾つかのカテゴリについての尺度を出力する。

- 1 ラティス全体の確率スコア (理想的な確率値は 1.0 であるが、枝刈りなどによりそうはならない。)
- 2 当該単語の事後確率スコア
- 3 競合候補のうち、音素系列が共通な部分パスをマージ (確率和) した、事後確率スコア
- 4 競合候補のうち、単語 (系列) が共通な部分パスをマージ (確率和) した、事後確率スコア
- 5 ターゲット単語区間に出現する、全ての部分パスによる各事後確率のエントロピー
- 6 ターゲット単語区間に出現する、全ての部分パスをその音素系列でマージ (確率和) した後の、エントロピー
- 7 ターゲット単語区間に出現する、全ての部分パスをその単語系列でマージ (確率和) した後の、エントロピー

ここでカテゴリとは、トータルスコア (音響 + 言語)、音響スコア、言語スコア、クラス Ngram スコア、単語出力スコアを指し、このそれぞれについて上記7つの値を出力する。出力されるカテゴリは、A.2.1節に述べたオプション `-ATRresult:FBProb=` に与えた数値により異なる。数値と出力されるカテゴリの関係は次のようになっている。

- FBProb=1 トータルスコアのみ
- FBProb=2 トータル, 音響, 言語 スコア
- FBProb=3 トータル, 音響, 言語, クラス Ngram, 単語出力スコア

A.5.16 CMPosteriori2

単語グラフによる当該単語の事後確率スコアから、競合候補間の事後確率スコア差を求める。(これはナンセンス。) このモジュールを利用するにも、`CMPosteriori.pm` 同様、ラティス中の各ノードにおける前向き確率と後向き確率を求めておく必要がある。また出力する尺度も、各カテゴリごとの次の7つの尺度となっている。

- 1 競合候補のうち最大事後確率スコアと、当該単語の事後確率スコアの差
- 2 競合候補のうち最大事後確率スコアと、当該単語と同一単語のうち最大の事後確率スコアの差
- 3 競合候補のうち最大事後確率スコアと、当該単語と同一音素系列の競合部分パスのうち、最大の事後確率スコアの差
- 4 当該単語とは単語が異なる競合候補のうち最大事後確率スコアと、当該単語の事後確率スコアの差
- 5 当該単語とは単語が異なる競合候補のうち最大事後確率スコアと、当該単語のうち最大の事後確率スコアの差
- 6 当該単語とは音素系列が異なる競合候補のうち最大事後確率スコアと、当該単語の事後確率スコアの差
- 7 当該単語とは音素系列が異なる競合候補のうち最大事後確率スコアと、当該単語と同一音素系列のうち最大の事後確率スコアの差

なお、カテゴリの種類と、出力カテゴリの定義方法は `CMPosteriori.pm` と同じである。

A.5.17 CMClassHypo

当該単語についてクラス Ngram 確率や単語出力確率を求め、さらに所属するクラスを出力する。 `CalcCM.pl` 出力の第6カラム以降の意味は次のようになる。

- 6カラム 先行単語のクラスから当該単語のクラスへのクラス 2gram 確率
- 7カラム 当該単語クラスからの単語出力確率
- 8カラム 当該単語のクラス名
- 9カラム “`CMStypHypo`” による第6カラムと同じ
- 10カラム “`CMStypHypo`” による第7カラムと同じ
- 11カラム “`CMStypHypo`” による第9カラムと同じ
- 12カラム 当該単語と競合する単語クラスの種類の数 (単語クラスマージ)

A.5.18 CMCompArc

当該単語と競合する各部分パスについて、この部分パスを通る最適な全体パスを探索、1st-Bestのパスとこの競合パスとのスコアの差などを求める。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

6カラム	当該単語区間で競合する部分パスの、単語系列の種類数
7カラム	当該単語を通る最適パスと、競合部分パスを通る最適パスのうち最大のもののトータルスコア(音響+言語)の差
8カラム	当該単語を通る最適パスと、競合部分パスを通る最適パスのうち最大のものの音響スコアの差
9カラム	当該単語を通る最適パスと、競合部分パスを通る最適パスのうち最大のものの言語スコアの差
10カラム	第8カラムの値を区間長(フレーム数)で正規化したもの
11カラム	第9カラムの値を区間長(単語数)で正規化したもの
12カラム	第10カラムと第11カラムの和
13カラム	当該単語区間で、ベストパスと競合2位パスを入れ換えるために必要なLMスケールの変動幅

A.5.19 CMPhDuration

当該単語を構成する各音素について、事前に求めておいた音素持続時間分布から持続時間確率を求める。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

6カラム	当該単語区間中で、各音素のうち最低の音素持続時間確率
7カラム	第6カラムの対数値
8カラム	当該単語区間中で、各音素の持続時間確率の対数値の平均
9カラム	当該単語区間中で、各音素の持続時間確率の平均

A.5.20 CM2gramCount

トレーニングデータ中に出現した、クラス2gramなどの言語現象の発生回数などを当該単語周辺について求める。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

6カラム	学習データ中に出現した先行単語クラス・当該単語クラス対の数
7カラム	当該単語の先行単語クラスから後続クラスへの遷移エントロピー
8カラム	“CMStypHypo”による第6カラムと同じ
9カラム	“CMStypHypo”による第7カラムと同じ
10カラム	“CMStypHypo”による第8カラムと同じ
11カラム	“CMStypHypo”による第9カラムと同じ

A.5.21 CMLangEntropy

当該単語について、その先行単語からのクラス遷移エントロピーおよび、単語出力確率のエントロピーを求める。CalcCM.pl出力の第6カラム以降の意味は次のようになる。

6カラム	学習データ中に出現した先行単語クラス・当該単語クラス対の数
7カラム	当該単語の先行単語クラスから後続クラスへの遷移エントロピー
8カラム	当該単語のクラスから後続クラスへの遷移エントロピー
9カラム	当該単語と競合する単語のうち、その先行単語のクラスの遷移エントロピーが最大のものの値
10カラム	当該単語と競合する単語に先行する単語クラスの、各遷移エントロピーの和
11カラム	当該単語と競合する単語に先行する単語クラスの、各遷移エントロピーの平均
11カラム	当該単語の属するクラスの、単語出力エントロピー
12カラム	当該単語の属する To クラスから先行の From クラスへの逆向きの遷移エントロピー
13カラム	注目単語の属する From クラスの単語出力エントロピー
14カラム	第6カラムの対数値