社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2006－SLP－64 (22)
2006／12／21

単語グラフを利用したベイズリスク最小化音声認識とそれに基づく重要文抽出

南條 浩輝† 河原 達也††

† 龍谷大学 理工学部 情報メディア学科
〒520-2194 大津市瀬田大江町横谷 1-5
†† 京都大学 情報学研究科 知能情報学専攻
〒606-8501 京都市左京区吉田二本松町
E-mail: †nanjo@ryukoku-u.jp

あらまし ドメインを限定しない自然な話し言葉の音声理解を目的とした音声認識とそれを用いた音声理解について述べる．音声理解を困難にする単語の認識誤りを重要視する評価尺度である「重みつき単語誤り率（weighted word error rate: WWER)」について述べ，その上で，その重みつき単語誤り率の最小化を行う音声認識，具体的にはベイズリスク最小化（Minimum Bayes-Risk: MBR）の枠組みに基づいた音声認識手法について述べる．これまでに N-best リストを用いた MBR 音声認識の効果を確認しているが，N-best リストは認識競合候補の表現力が弱く，MBR 音声認識によるリスク最小化の効果が小さいという問題があった．このため，本稿では，N-best リストよりも表現力の大きい単語グラフを利用した MBR 音声認識に焦点をあてる．はじめに種々の単語グラフ探索アルゴリズムの比較を行い，コンフュージョンネットワーク型の単語グラフ探索の有効性を示す．次に CSJ の学会講演 10 講演を用いて音声認識と音声理解のファーストステップである音声の重要箇所の同定（重要文抽出）の評価を行い，提案する手法が音声認識および重要文抽出の両方に対して効果があることを示す．

キーワード 音声認識，音声理解，ベイズリスク最小化デコーディング，重要文抽出，単語グラフ

Minimum Bayes-Risk Decoding using Wordgraph and Key-sentence Extraction

Hiroaki NANJO† and Tatsuya KAWAHARA††

† Faculty of Science and Technology, Ryukoku University
Seta, Otsu 520-2194 Japan
†† Graduate School of Informatics, Kyoto University,
Sakyo-ku, Kyoto 606-8501 Japan
E-mail: †nanjo@ryukoku-u.jp

Abstract Speech recognition strategy for open-domain speech understanding and its application to a key-sentence extraction, which is a first step of speech understanding, are presented. Firstly, a weighted word error rate (WWER), which is an evaluation measure giving a weight on errors from a viewpoint of speech understanding, is described. Then, we show a decoding method to minimize WWER based on Minimum Bayes-Risk (MBR) framework. We have investigated N-best list based MBR decoding. In this paper, several wordgraph based MBR decoding algorithms are presented, and then, we show the effectiveness of confusion network based decoding strategy. Finally, we demonstrate that the decoding method works reasonably for improving both WWER and key sentence indexing.

Key words speech recognition, speech understanding, Minimum Bayes-Risk decoding key sentence extraction, wordgraph

1. はじめに

近年，大規模な『日本語話し言葉コーパス』(CSJ: Corpus of Spontaneous Japanese) [1] [2] が整備され，話し言葉の音声認識および音声理解の研究がさかんに行われるようになってきた．我々も CSJ を対象とした音声認識 [3]，および音声の内容理解を目的とした音声の重要箇所を同定する研究（重要文抽出）を行ってきた [4] [5]．本稿では，それらを個別に扱うのではなく，音声理解システムのフロントエンドとしての音声認識手法とそれを用いた重要文抽出について述べる．

音声理解を行う際には，理解に必要な語（=キーワード）や構文情報などの高度な知識を用いることが多く [6] [7]，実際に我々の音声理解（重要文抽出）の研究でもそのようなアプローチを採用している [5]．このような情報を用いる必要がある音声理解を目的とした場合には，音声認識時に全ての単語を一様に認識するのではなく，キーワードや構文解析などに必要な単語を優先的に認識するのが望ましい．これまでの話し言葉の音声認識は，発話全体の忠実なテキスト化を目的とするものが主であり，音声認識の精度評価は単語誤り率（WER）に基づいて行われるのが一般的であった．しかし，この単語誤り率による評価では，音声理解に影響の大きい語（キーワードなど）とそうでない語（間投詞やつなぎ語など）の認識誤りが同等に扱われており，音声理解の観点からは音声認識の評価尺度として必ずしも最適ではない．

キーワードの誤り率を音声認識の評価尺度として用いる研究もみられるが，これらは主にタスク限定の対話システムを対象としており，内容理解に必要なキーワードリストを事前に定義できることが前提である．これに対して，本研究では，講演や講義のような“ドメイン非限定”の発話を対象としており，事前にキーワードリストを用意することは困難である．したがって，従来の単純なキーワード認識・評価の枠組をそのまま適用することはできない．

これに対し，我々は講演音声の理解を目的とした音声認識の評価尺度と認識手法を提案してきた [8]．具体的には，音声認識の評価尺度として情報検索の観点から各単語の認識誤りに異なる重みを与える「重みつき単語誤り率（Weighted Word Error Rate: WWER）」を，認識手法としてベイズリスク最小化（Minimum Bayes-Risk: MBR）デコーディングの枠組に基づいた手法を提案してきた．この概要を図 1 に示す．実際に N-best リストを用いたベイズリスク最小化音声認識を行っており，その効果を確認している [8] [9]．ただし，N-best リストは認識競合候補の表現力が弱く，MBR 音声認識によるリスク最小化の効果が小さいという問題があった．

このような背景に基づいて，本稿ではまず，N-best リストよりも表現力の大きい単語グラフを用いたベイズリスク最小化音声認識アルゴリズムについて検討を行う．具体的には，音声認識デコーダ Julius が出力する単語グラフの利用法に関して種々の検討を行い，コンフュージョンネットワーク型の単語グラフ探索の有効性を示す．次に，単語グラフ探索を用いたベイズリスク最小化音声認識を利用して重要文抽出を行い，提案手法が音声認識と重要文抽出の両方に効果があることを示す．

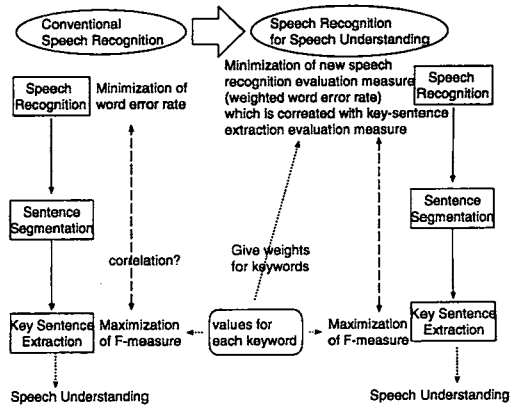

図 1　音声理解を目的とした音声認識

Fig. 1　Speech Recognition for Speech Understanding

2. 音声理解のための音声認識の評価尺度

2.1 単語誤り率

音声認識の評価尺度で最も一般的なものは単語誤り率（WER）であり，これは式（1）で定義されるものである．

$$\mathrm{WER} = \frac{I+D+S}{N} * 100 \tag{1}$$

ここで，N は正解文における単語の数，S は置換誤り単語の数，D は削除誤り単語の数，I は挿入誤り単語の数である．

式（1）からも明らかなように，この評価尺度（単語誤り率）は，全ての単語の音声認識誤りを同一に扱っている．すなわち，単語誤り率は音声理解に決定的に影響を及ぼす重要な単語（キーワードのような単語）とそうでない単語（間投詞やフィラーなど）を区別せず，等しく扱う評価尺度である．

2.2 重みつき単語誤り率 – 単語誤り率の一般化 –

重要単語の認識誤りは音声理解に影響が大きいため，そのような誤りに重みをつけて評価する必要がある．本節では，各単語の音声認識誤りに異る重みを与えるために，単語誤り率（WER）の一般化を行なう．具体的には，以下の式で定義する「重みつき単語誤り率（WWER）」について述べる．

$$\mathrm{WWER} = \frac{V_I + V_D + V_S}{V_N} * 100 \tag{2}$$

$$V_N = \Sigma_{w_i} \quad v_{w_i} \tag{3}$$

$$V_I = \Sigma_{\hat{w}_i \in I} \quad v_{\hat{w}_i} \tag{4}$$

$$V_D = \Sigma_{w_i \in D} \quad v_{w_i} \tag{5}$$

$$V_S = \Sigma_{seg_j \in S} \quad v_{seg_j} \tag{6}$$

$$v_{seg_j} = \max(\Sigma_{\hat{w}_i \in seg_j} v_{\hat{w}_i}, \Sigma_{w_i \in seg_j} v_{w_i}) \tag{7}$$

ここで，v_{w_i} は正解文における i 番目の単語（w_i）の重みを表し，$v_{\hat{w}_i}$ は音声認識結果における i 番目の単語（$\hat{w}_i$）の重みを表している．また，seg_j は j 番目の置換誤り区間を表し，

ASR result	:	a	b	c	d	e	f	
Correct transcript	:	a		c	d'		f	g
DP result	:	C	I	C	S		C	D

C: correct word, I: inserted word,
D: deleted word, S: substituted segment

$$\mathrm{WWER} = (V_I + V_D + V_S)/V_N * 100$$

$$V_N = v_a + v_c + v_{d'} + v_f + v_g$$
$$V_I = v_b$$
$$V_D = v_g$$
$$V_S = \max(v_d + v_e, v_{d'})$$

v_i: weight of word i.

図 2 重みつき単語誤り率（WWER）の計算例
Fig. 2 Example of weighted word error rate (WWER) calculation

v_{seg_j} は誤り区間 seg_j の重みを表している．この誤り区間の重み v_{seg_j} は，当該区間 seg_j に含まれる正解系列の単語の重みの合計値と認識結果の単語の重みの合計値のうち，大きいほうとする．参考のため，図 2 に重みつき単語誤り率の計算例を示す．

なお本研究では，誤り単語を同定する際に，単語誤り率を求める際に用いた DP マッチングの結果を用いている．したがって，全ての単語の重みを等しく設定した場合，重みつき単語誤り率は単語誤り率と等しくなる．

3. ベイズリスク最小化デコーディング

前章では，音声理解のための音声認識の評価尺度である重みつき単語誤り率（WWER）の定式化を行なった．本章では，この重みつき単語誤り率を最小化する音声認識手法（デコーディング手法）について述べる．本手法は，ベイズリスク最小化枠組み [10] に基づく音声認識手法である．

3.1 ベイズリスク最小化音声認識の枠組み

統計的な音声認識は一般的に，与えられた入力音声信号 X を最もよく説明する単語列 $\hat{W}$ を求めるプロセスとして定式化される．これを式 (8) に示す．

$$\hat{W} = \underset{W'}{\mathrm{argmax}}\, P(W'|X) \tag{8}$$

ベイズ決定理論に基づくと，音声認識は決定規則（$\delta(X) : X \rightarrow \hat{W}$）と記述できる．ここで，損失関数を $l(W, \delta(X)) = l(W, W')$ とすると，音声認識は以下のベイズリスク最小化の枠組みで記述できる [10]．

$$\delta(X) = \underset{W}{\mathrm{argmin}} \sum_{W'} l(W, W') \cdot P(W'|X) \tag{9}$$

式 (9) の右辺の事後確率はベイズ則を用いて $P(W', X)/P(X)$ と展開でき，分母の $P(X)$ は式全体の最小化に影響を与えないため省略できる．さらに，それぞれのスコアに重みパラメータを乗じた式 (10) を使う有効性が先行研究で示されており [10]，本研究でもこれを導入する．重みパラメータの値はディベロプメントセットを用いて推定する．

$$\delta(X) = \underset{W}{\mathrm{argmin}} \sum_{W'} l(W, W')^{\lambda_1} \quad P(W', X)^{\lambda_2} \tag{10}$$

なお，式 (8) で示されている一般的な音声認識のプロセスは，式 (9)（10）において 0/1 損失関数を用いる場合と等価であり，文誤り率（Sentence Error Rate）を最小化するプロセスであるといえる．

単語誤り率最小化を目的とした場合は，損失関数 $l(W, W')$ として単語誤り率（もしくは最小編集距離=Levenshtein distance）を用いればよいことが知られている [10] [11]．本研究では，重みつき単語誤り率の最小化を目的とするので，損失関数 $l(W, W')$ として重みつき単語誤り率を用いればよい．

4. 単語グラフ探索アルゴリズム

次に，単語グラフ（WG）を利用したベイズリスク最小化音声認識の手法について述べる．本研究では，単語グラフを生成するために音声認識エンジン Julius rev.3.5 を用いる．具体的には，Julius の第 2 パス中に動的生成された単語グラフを用いる．

4.1 単語グラフに基づく音声認識

本研究での単語グラフ探索アルゴリズムは以下の通りである．

（1） 仮説スタックにグラフ上の単語のうちで左側に単語を持たないものを初期仮説としてスタックに積む．

（2） 仮説スタックの全ての仮説の評価値を計算する．

（3） 仮説スタックの全て仮説が完全パスになるまで 4, 5 を繰り返す．

（4） 仮説スタック中の仮説の中で評価値の最も小さい仮説を取り出し，それを展開して得られる仮説を全て仮説スタックに積む．

（5） 仮説スタックの全ての仮説の評価値を計算する．仮説数が一定値を超えた場合は，評価値の高い仮説を枝刈りする．

（6） 仮説スタックの中で最も評価値の小さい仮説を結果として出力する．

なお，仮説スタック中の仮説 W_i の評価値 $f(W_i)$ の計算は式 (11) に基づいて行う．

$$f(W_i) = \sum_{W_j \in stack} l(W_j, W_i) \cdot P(W_j) \tag{11}$$

where

$$P(W_j) = P(W_j, X(W_j)) \cdot (\sum_{W_{e_j}} P(W_{e_j}, X(W_{e_j})))$$

$$l(W_j, W_i) = \min l(W_j W_{e_j}, W_i W_{e_i})$$

ここで，$X(W_j)$ は仮説 W_j の存在時刻に対応する観測音声信号であり，W_{e_j} は仮説 W_j から展開できるパスを表している．したがって，$P(W_j)$ は，これまで展開した仮説の音声認識スコアとこれから展開できるパスの認識スコアの合計となる．損失 $l(W_j, W_i)$ は，仮説 W_j を展開して得られる仮説集合と仮説 W_i を展開して得られる仮説集合のそれぞれの仮説の間での損失の最小値である．

4.2 N-best リストに基づく音声認識

N-best リストを用いたベイズリスク最小化音声認識アルゴリズムは以下の通りである．

（1） 音響モデルと言語モデルを用いて単語グラフの探索を行い，音声認識スコアの高い順にパスを N 個求め，N-best リストを作成する．

（2） 各パスの評価値を計算する．

（3） 評価値の最も小さな仮説を結果として出力する

なお，仮説 W_i の評価値 $f(W_i)$ は以下の式に基づいて計算する．

$$f(W_i) = \sum_{W_j \in \text{N-best list}} l(W_j, W_i) \cdot P(W_j) \qquad (12)$$

ここで，$P(W_j)$ は仮説 W_j の認識スコア，$l(W_j, W_i)$ は仮説 W_j を W_i と誤ったときの損失である．N-best リスト探索では仮説 W_i は全て完全パスとなり，式（11）と式（12）の本質的な違いはその点のみである．

4.3 コンフュージョンネットワークに基づく音声認識

最後に，単語グラフそのもの（WG）を探索するのではなく，それを圧縮したグラフ，具体的には図 3 に示すコンフュージョンネットワーク（CN）の探索について説明する．

本研究での CN の作成は次の手順で行う．

（1） 単語グラフ中の最ゆうパスを求める．

（2） 最ゆうパスの各単語ごとに，競合する単語（競合候補）を単語グラフから求めて競合候補セットを生成する．競合候補の判定は存在時刻に基づいて行う．

（3） 競合候補セットごとに各単語の事後確率を計算する．

（4） 競合候補セットごとに事後確率がしきい値以下の単語を削除し，それらの事後確率の合計を NULL リンクの事後確率として設定する．

ここで，単語の事後確率は，当該単語を含む仮説の最ゆう値を用いて計算する．具体的には，各単語のスコアを，当該単語を含む仮説の最ゆう値と単語グラフ上で当該単語を通るパスの数とを乗じたものと定義し，式（13）に基づいて事後確率化を行う．なお，競合候補セット中に同一単語があれば，それらを単語のスコアも含めてマージしたのちに，式（13）に基づいて事後確率を求める．

$$\frac{P(w_i)^\alpha}{\sum_{j=1}^{N} P(w_j)^\alpha} \qquad (13)$$

式（13）の $P(w_i)$ はある競合候補セット中の i 候補目の単語のスコアを表し，N は当該競合候補セットに含まれる単語数を示すものとする．α はスケーリングパラメータであり，本研究では 0.05 とした．

次に CN 探索について述べる．一般的には，CN 上の各競合候補セットごとに事後確率が最大の単語を選択すれば単語誤り率最小化音声認識となる [12]．単語グラフを CN として表現した場合，前後の接続関係の制約がなくなるという懸念があるが，事後確率の計算時に N-gram 言語モデルや triphone モデルを用いることである程度モデル化できると考えられている．

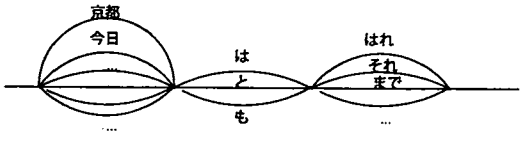

図 3 コンフュージョンネットワーク
Fig. 3 Confusion Network

表 1 ベイズリスク最小化音声認識-探索アルゴリズムの比較
Table 1 MBR decoding – comparison of search algorithms

音声認識	単語誤り率（WER）
SER 最小化（ゆう度最大化）	30.6%
WER 最小化（WG）	31.5%
WER 最小化（N-best リスト）	30.1%
WER 最小化（CN）	28.5%

一方，本手法では，ある単語の事後確率を求める際に，当該単語を通過する最ゆうパスの音声認識スコアを単純に当該単語を通過するパス数と乗じているため，厳密な事後確率の計算がなされていない．この結果，特に言語制約が弱められていると考えられる．したがって，本研究では CN 探索を次のように行う．

（1） 事後確率と言語モデルゆう度を用いて CN の探索を行い，仮説スコアの高い順にパスを N 個求め，N-best リストを作成する．仮説スコアは，当該仮説に含まれる単語の事後確率の累積値と音声認識で用いた言語モデル（3gram）によって計算される言語ゆう度を混合して定義する．本研究では混合比は事後的に求めた（混合比は 2:1）．

（2） N-best リストを 4.2 節で述べた手法（式（12））で表されるリスク最小化音声認識を行い，評価値の最も小さい仮説を出力する．

4.4 ベイズリスク最小化音声認識の評価

単語グラフ（WG），N-best リスト，コンフュージョンネットワーク（CN）を用いてベイズリスク最小化音声認識を行い，各探索手法の比較を行った．ここでは，最小編集距離を損失関数とし，単語誤り率最小化で評価を行った．評価には CSJ のテストセット 1（学会講演 10 講演）を用いた．単語グラフが構築できない発話がいくつか存在したが，それらについては Julius の第 2 パスの結果（1-best）をそのまま出力した．

結果を表 1 に示す．CN を用いたベイズリスク最小化デコーディングで最も高い認識率が得られ，ベースラインに比べて 6.9%の単語誤り率の削減（絶対値で 2.1%）が得られた．N-best リストを用いた場合は，ベースラインに比べて 1.6%の単語誤り率の削減となった．一方，単語グラフそのものを用いて探索を行った場合は認識率の改善は見られなかった．この主な原因は，グラフサイズが大きく探索途中で最ゆう仮説が枝刈りされたためと考えられる．

次に CN 探索時の言語モデル制約の効果について述べる．CN 探索時の各仮説のスコア計算方法の比較を表 2 に示す．従来のように事後確率のみに基づくスコア計算では，認識率の改善が得られなかった．主な理由は，言語制約が弱められているためと考えられる．実際にこの手法で出力された結果には言語的に不自然なものが多く観察された．言語モデルのみを用いて仮説

表 2　CN 探索時の仮説スコア計算方法の比較

Table 2　Comparison of several calculation methods of hypothesis score

スコア計算	単語誤り率（WER）
事後確率	30.7%
言語モデル	32.2%
事後確率+言語モデル	28.5%

※ WER 最小化音声認識

スコアを決定しても効果は見られなかったが，事後確率スコアと言語モデルスコアの両方を利用して仮説スコアを決定することで約 7%の誤り率削減を実現できた．誤り率の削減はテストセットの講演全てで確認できた．

5. ベイズリスク最小化音声認識を用いた講演からの重要文抽出

本章では，重要文抽出における重みつき単語誤り率最小化音声認識の有効性について述べる．

5.1 重要文抽出の評価データ及び評価方法

CSJ の一部の講演，具体的にはコアに含まれる講演および音声認識のテストセットの講演，には書き起こしに重要文タグ（講演ごとに 3 名，全文の 50%及び 10%を目安にタグ付け）が付与されている．本研究では，4. 章で MBR 音声認識の評価に用いたものと同じもの（CSJ 音声認識のテストセット 1 の 10 講演: すべて学会講演）を評価に用いる．

次に，評価を行う際に正解とする重要文の設定と評価法について述べる．重要文として抽出される文は 3 名の作業者で異なるため，正解とする重要文は，3 名から任意の 2 名を選び，その 2 名がともに 50%の重要文抽出において抽出した文とする．評価は，3 通りの正解それぞれに対する再現率（recall），適合率（precision），F 値を算出し，それらの平均で行う．ここで，F 値は再現率と適合率の調和平均であり，式（14）により定義される．

$$\text{F-measure} = \frac{2 * recall * precision}{recall + precision} \tag{14}$$

この 2 名一致による正解重要文に対して，他の 1 名が抽出したものを照合することで，人間の重要文抽出精度が算出できる．人間の重要文抽出精度は，再現率 82.2%，適合率 60.6%，F 値 0.698 であった．これらは，重要文抽出システムの評価の際に，目標となる数値である．

5.2 講演の談話標識を利用した重要文抽出

次に，本研究で用いた重要文抽出手法について述べる．

講演には，論理的に順序づけられたいくつかの論点（話題）があり，話題の転換点には重要な発話が多い．我々の重要文抽出の手法はこの点に着目し，話題の転換点に頻出する語（談話標識）の統計量，すなわち単語頻度と文頻度に基づく統計量 DM_m $(= wf_m * \log(N_s/sf_m))$ を用いてセクション境界の検出と重要文の抽出を行うものである [5]．ここで wf_m，sf_m は，それぞれ話題の転換点の冒頭（文集合）において単語 m が出現する回数，全ての文（総数 N_s）で単語 m が出現する文数である．話題境界に頻出し，その他ではあまり出現しない単語に対して，この値 DM は大きな値となる．各文に出現する異なる全ての談話標識の統計値の合計 S_j^{DM} $(= \sum_i DM_{m_i})$ は話題境界らしさ（=重要度）を表すので，この評価値が高いものを重要文として抽出する．

表 3　学会講演から抽出された談話標識

Table 3　Presumed discourse markers from academic oral presentations

次	先程	今	実際	今回	最初	最後
研究	説明	発表				
結果	実験	問題	評価	方法	目的	

表 4　50%重要文抽出結果

Table 4　Result of extraction of key sentences by 50%

重要文抽出	文分割	再現率	適合率	F 値
人間	人手	82.2%	60.6%	0.698
自動	人手	69.9%	51.5%	0.593
自動	自動	71.4%	46.9%	0.566

学会講演から単語ごとに統計量 DM_m を求めた．大きい DM 値が得られた単語を表 3 に示す．最上段の単語は，一般的な談話標識と一致する．二段目以降の単語は講演スタイルに多少依存しているものの，新しい話題を提示する際に頻繁に用いられる単語である．

談話標識を用いた重要文抽出の結果を表 4 に示す（中段）．再現率，適合率，F 値はそれぞれ，69.9%，51.5%，0.593 であり，人間の精度の約 85%であることがわかる．

次に，重要文抽出における文分割の影響について述べる．重要文抽出を行うためには講演の書き起こしを文に分割する必要があるが，話し言葉では文の単位が明確でない（単純にポーズで区切ることが不可能）ため，問題となる．我々は自動文分割の研究 [13] を行っており，本研究でもこれを用いた．結果は表 4 の最下段に示されている．適合率が低下したものの再現率の低下はみられなかった．自動文分割を用いた場合でも，本手法は重要文抽出，特に重要文のインデキシングの手法として有用であることがわかる．

5.3 ベイズリスク最小化音声認識の効果

5.3.1 単語の重みの設定

前説でも述べたように，我々の重要文抽出システムでは談話標識の統計量を用いている．したがって本研究では，音声認識時の単語重みとして談話標識にはこれらの統計量を与え，それ以外の単語には重み 0 を与える．重みつき単語誤り率（WWER）はこの重みを式 (2) に適用して算出する．

5.3.2 実験結果

重みつき単語誤り率（WWER）最小化音声認識を行うため，損失関数を重みつき最小編集距離（式 (2) の分子）として定義し，ベイズリスク最小化音声認識を行った．探索手法には，最も効果の高かったコンフュージョンネットワーク（CN）探索を用いた．結果を表 5 に示す．比較のために，WER 最小化音声認識（損失関数を最小編集距離としたもの）の結果も示してい

表 5　重みつき単語誤り率（WWER）最小化デコーディングの効果
Table 5　Results of WWER minimizing decoding

全講演（10 講演）

minimization target (loss function)	WER	WWER	key sentence indexing accuracy (F-measure)
1/0 – baseline	30.6	23.0	0.553
WER	**28.5**	23.8	0.554
WWER	29.2	**22.7**	**0.555**
manual transcription	0	0	0.566
manual transcription2	0	0	0.593

manual transcription2: 文分割も人手で実行

WWER 改善が得られた 5 講演

minimization target (loss function)	WER	WWER	key sentence indexing accuracy (F-measure)
1/0 – baseline	37.5	32.7	0.550
WER	**35.6**	33.6	0.552
WWER	36.2	**31.6**	**0.555**
manual transcription	0	0	0.557
manual transcription2	0	0	0.585

manual transcription2: 文分割も人手で実行

WWER 改善が得られなかった 5 講演

minimization target (loss function)	WER	WWER	key sentence indexing accuracy (F-measure)
1/0 – baseline	23.1	**13.7**	**0.557**
WER	**20.9**	14.5	0.555
WWER	21.5	14.3	0.555
manual transcription	0	0	0.575
manual transcription2	0	0	0.600

manual transcription2: 文分割も人手で実行

る．ベイズリスク最小化音声認識を行うことで，それぞれ対象とする評価尺度を改善できていることがわかる．

WWER 最小化音声認識により，10 講演の平均で WWER が 23.0%から 22.7%に改善され，それに伴い，重要文抽出精度（F 値）も 0.553 から 0.555 に向上した．効果が小さかったため分析を行ったところ，WER 最小化音声認識では全ての講演で WER の改善が確認できていたのに対して，WWER 最小化音声認識では半分の講演（5 講演）でのみ WWER が改善できていた．表 5 には，WWER の改善が得られた講演とそうでない講演にわけて評価した結果も示している．比較的認識率が低い講演で 3.4%の WWER の削減（絶対値では 1.1%）が得られており，この WWER の改善に伴って重要文抽出の精度も 0.550 から 0.555（人手による書き起こしを用いた場合は 0.557）に向上していることがわかる．一方，WWER の改善がみられなかった講演については，重要文抽出の精度の向上も見られなかった．WER 最小化音声認識を行った場合は，音声認識の精度向上に伴う重要文抽出の精度向上はほとんどみられなかった．これらのことは，WWER 最小化音声認識が重要文抽出（音声理解）に効果があることを示している．なお，WWER の改善がみられなかった講演については，その原因の分析を今後行っていく必要がある．

最後に，文分割の影響について述べる．表 4 および表 5 の各表の下段には，音声認識誤りが 0%の場合の重要文抽出精度を示している．自動文分割による重要文抽出精度の低下が非常に大きい（0.593→0.566）ことがわかる．この結果は，自動文分割システムの精度を向上させる重要性，および自動文分割に影響を及ぼす単語に重みをつけてベイズリスク最小化音声認識を行う必要性を示唆している．

6. おわりに

ドメインを限定しない自然な話し言葉の音声理解を目的として単語の重要度を考慮したベイズリスク最小化枠組に基づく音声認識とそれを用いた重要文抽出の研究を行った．CSJ の講演を対象として種々のベイズリスク最小化音声認識アルゴリズムを比較を行い，コンフュージョンネットワークを用いたベイズリスク最小化音声認識の有効性を示した．次に，重要文抽出に影響の大きい単語を重視する音声認識の評価尺度「重みつき単語誤り率」を定義し，これを最小化するベイズリスク最小化音声認識を行った．その結果を用いて重要文抽出を行い，提案手法が重要文抽出（音声理解）に効果があることを示した．

文　献

[1] 前川喜久雄．言語研究における自発音声．音講論, 1-3-10, 春季 2001.

[2] 小磯花絵, 前川喜久雄．『日本語話し言葉コーパス』の概要と書き起こし基準について．情処学研報, 2001-SLP-36-1, 2001.

[3] H.Nanjo and T.Kawahara. Language model and speaking rate adaptation for spontaneous presentation speech recognition. *IEEE Trans. Speech & Audio Process.*, Vol. 12, No. 4, pp. 391–400, 2004.

[4] 南條浩輝, 北出祐, 河原達也．談話標識の統計的選択に基づいた CSJ の講演からの重要文抽出．電子情報通信学会技術研究報告, SP2003-125, NLC2003-62 (SLP-49-13), 2003.

[5] T.Kawahara, M.Hasegawa, K.Shitaoka, T.Kitade, and H.Nanjo. Automatic indexing of lecture presentations using unsupervised learning of presumed discourse markers. *IEEE Trans. Speech & Audio Process.*, Vol. 12, No. 4, pp. 409–419, 2004.

[6] 野畑周, 関根聡, 内元清貴, 井佐原均．話し言葉コーパスにおける文の切り分けと重要文抽出．「話し言葉の科学と工学」ワークショップ予稿集, pp. 93–100, 2002.

[7] 菊池智紀, 古井貞熙, 堀智織．重要文抽出と文圧縮による音声自動要約．信学技報, SP2002-158, 2002.

[8] 南條浩輝, 河原達也．音声理解のための音声認識評価尺度とベイズリスク最小化デコーディング．電子情報通信学会技術研究報告, SP2004-122, NLC2004-82 (SLP-54-42), 2004.

[9] 南條浩輝, 翠輝久, 河原達也．単語の重要度を考慮したベイズリスク最小化音声認識とそれに基づく情報検索．電子情報通信学会技術研究報告, SP2005-99, NLC2005-66 (SLP-59-10), 2005.

[10] V.Goel, W.Byrne, and S.Khudanpur. LVCSR rescoring with modified loss functions: A decision theoretic perspective. In *Proc. IEEE-ICASSP*, Vol. 1, pp. 425–428, 1998.

[11] A.Stolcke, Y.Konig, and M.Weintraub. Explicit word error minimization in N-best list rescoreing. In *Proc. EUROSPEECH*, pp. 163–165, 1997.

[12] Y.Qian, F.K.Soong, and T.Lee. Tone-enhanced generalized character posterior probability(GCPP) for cantonese lvcsr. In *Proc. IEEE-ICASSP*, Vol. 1, pp. 133–136, 2006.

[13] 南條浩輝, 下岡和也, 河原達也．講演からの重要文抽出のための文区切り手法の改善．音講論, 3-Q-28, 春季 2004.