Vol.2013-SLP-97 No.7
2013/7/26

ベイズリスク最小化音声認識の複数仮説を用いた音声検索

南條 浩輝[1,a)]　古谷 遼[2]

概要：音声検索（音声入力型情報検索）の研究を行う．音声検索システムは，音声認識結果から検索要求を生成して情報検索を行うものである．これまで音声認識に関しては，ベイズリスク最小化の枠組みに基づいて検索にとって重要な単語の誤りを少なくする手法が提案されている．また，音声認識結果からの検索要求生成においては，認識結果の複数仮説を用いて検索要求を生成する研究がこれまでに行われている．しかし，これらの組み合わせ，すなわちベイズリスク最小化音声認識の結果からの検索要求の生成において，認識結果の複数仮説を利用する方法はこれまでにみられない．このような背景に基づき，本研究ではこれらの組み合わせ，すなわち，重要単語の誤りを最小化する音声認識を行って N-best 候補を出力し，検索要求を生成することを提案し，その評価を行う．日本語講演音声ドキュメントの検索タスクにおいて音声検索の評価を行ったところ，提案法が効果的であることがわかった．

1. はじめに

音声認識をフロントエンドに持つ情報検索，すなわち音声入力型の情報検索（音声検索）システムについて研究を行う [1], [2]．音声検索システムにおいては，バックエンドの検索システムが高精度であっても，音声認識で認識誤りが発生した場合，その影響を受けて検索精度が低下する．したがって，音声認識誤りによる影響を受けにくい音声検索を実現するための方式が求められている．

情報検索には，検索の観点から重要な語句（例えばキーワード）とそうでない語句が存在する．このため，音声検索における音声認識では全ての語句を同等に扱うことは適切でなく，重要語句の音声認識誤りを少なくすることが重要である．音声認識の評価は重要語句がどの程度認識されたかという観点で行う必要がある．このような研究の一つに，情報検索のキーワードの音声認識誤り，すなわちキーワード誤り率に着目して検索の性能向上を狙う方法がある [3]．この方法は，あらかじめキーワード集合を定義でき，かつ各キーワード間には差が無いタスクには十分である．しかし，キーワード間に差がある場合，例えばキーワードの重みを用いるベクトル空間モデルに基づく検索システムなどでは，全てのキーワードを同等に扱うことが適切とは限らず，誤ると影響が大きいキーワードを優先的に音声認識することが重要である．この問題に対して，検索における重要単語がどの程度誤っているかを測る指標として重み付き単語誤り率（WWER: Weighted Word Error Rate）が提案されており，さらにこれを削減するベイズリスク最小化（MBR: Minimum Bayes-Risk）音声認識が情報検索システムの性能向上に効果的であることが示されている [4]．しかし，MBR 音声認識の結果から検索要求をどのように生成するかに関する研究は不十分である．実際に，従来の MBR 音声認識に基づく音声検索では，単に認識結果の最上位仮説をそのまま用いている [4], [5]．

これらの背景に基づき，本論文では MBR 音声認識を用いた情報検索について研究を行う．具体的には，1) 重要単語の誤りを最小化する MBR 音声認識（WWER 最小化）を行い，2) リスクの小さい順に並び替えられた N-best リストから検索要求を生成する．このように，重要単語の誤り最小化を目的とする MBR 音声認識と，その結果からの検索要求生成手法の両方を組み合わせてその効果を調査した研究はこれまでになく，新しい．

本論文の構成について述べる．2 章で WWER および WWER 最小化音声認識について述べる．3 章で N-best リストを用いた検索要求生成手法について述べる．4 章で本論文での情報検索の評価尺度について述べ，5 章で提案手法が音声検索において有効であることを示す．6 章で結論を述べる．

[1] 龍谷大学理工学部
Faculty of Science and Technology, Ryukoku, University
[2] 龍谷大学理工学研究科
Graduate School of Science and Technology, Ryukoku University
[a)] nanjo@rins.ryukoku.ac.jp

2. 情報検索のためのベイズリスク最小化音声認識

2.1 重み付き単語誤り率

音声認識誤りにより検索性能を大きく低下させる語とそうでない語があるような音声検索システムでは，音声認識の評価尺度として各語の重要度を考慮した尺度が必要である．この評価尺度として，重み付き単語誤り率（WWER: Weighted Word Error Rate）がある [4]（式（1））．

$$WWER = \frac{V_I + V_D + V_S}{V_N} \tag{1}$$

ここで，V_I は挿入誤り単語の重要度の合計を，V_D は削除誤り単語の重要度の合計を，V_S は置換誤り区間の単語重要度の合計を，V_N は正解文の単語重要度の合計を表す．なお，誤り単語を同定する際には，単語誤り率（WER）を求める際と同様に DP マッチングの結果を用いる．全ての単語の重みを等しく設定したときには WWER は WER と一致し，WWER は WER を一般化したものとなっている．

2.2 ベイズリスク最小化音声認識

検索にとって重要な語に重みを与えて WWER を定義し，それを最小化するように音声認識を行うことで音声検索の精度向上が期待できる．このような音声認識はベイズリスク最小化（MBR: Minimum Bayes-Risk）の枠組み（式（2））[6], [7] で行うことができる [4].

$$\hat{W} = \arg\min_{W} \sum_{W'} l\left(W, W'\right)^{\lambda_1} P\left(W', X\right)^{\lambda_2} \tag{2}$$

ここで，$l\left(W, W'\right)$ は仮説 W' を仮説 W に誤った際の損失を求める損失関数を表し，$P\left(W', X\right)$ は入力信号 X と仮説 W' の同時確率（音声認識スコア）を表す．λ_1，λ_2 は損失関数および確率の重みパラメータである．損失関数として WER の分子に相当する編集距離（Levenshtein Distance）を用いると WER を削減する音声認識を行える [6], [7]．同様に，損失関数として WWER の定義式（式（1））の分子を用いると WWER を削減する音声認識を行える [4].

3. 音声認識結果からの検索要求の生成

3.1 ベクトル空間モデルに基づく情報検索システム

本論文では，情報検索システムとして一般的に広く用いられているベクトル空間モデルに基づくシステムを採用し，音声検索システムを構築する．ベクトル空間モデルでは，検索要求と文書をベクトルとして表現し，ベクトル間の類似度に基づいて検索を行う．ベクトルの要素には各索引語の出現頻度に基づく値を与える，すなわち各索引語に異なる重要度を与えることが一般的である．本論文でもそのようなシステムを構築する．

3.2 検索要求ベクトルの生成方法

音声検索システムにおいて入力音声から検索要求ベクトルを生成する方法として最も一般的な方法の一つに，音声認識を行ってテキストを生成し，そのテキスト中に含まれる索引語 t の出現数 qtf_t をベクトルの要素とする方法がある．音声認識時の最有力仮説（1-best 仮説）を使うのが一般的であるが，音声認識誤りによって 1-best 仮説に正しく索引語が含まれないことがある．この問題への対応として複数仮説の集合（N-best リスト）を用いること [8], [9] が考えられる．本論文では，N-best リストを用いて検索要求ベクトルを生成する手法を研究する．その際に MBR 音声認識を適用して N-best リストを作ることで検索性能の向上につながる検索要求ベクトルの生成を実現する．

3.3 複数仮説からの検索要求ベクトルの生成

3.3.1 仮説の順位に基づく重みを与える仮説統合

音声認識結果の N-best リスト全体をテキストとみなして索引語 t の出現数 qtf_t を求める．ただし，通常 N-best リストの上位仮説ほど認識精度が高く，上位の仮説に含まれる単語ほど正解単語である可能性が高い．このことから，検索要求中での索引語 t の出現数 qtf_t をそれが出現する仮説の順位に基づく重みを用いて調整する．本論文では順位に基づく重みとして以下の 3 つを採用し，3 種類の方法を実現する．

- N-best（一様）：順位によらず一様な重みを与える方法
- N-best（線形）：順位の逆数を重みとして与える方法
- N-best（対数）：順位の対数の逆数を重みとして与える方法

N-best（一様），N-best（線形），N-best（対数）はそれぞれ式（3），式（4），式（5）で qtf_t を計算する．

$$\mathrm{qtf}_t = \sum_{n=1}^{N} \mathrm{qtf}_{t,n} \tag{3}$$

$$\mathrm{qtf}_t = \sum_{n=1}^{N} \frac{\mathrm{qtf}_{t,n}}{n} \tag{4}$$

$$\mathrm{qtf}_t = \sum_{n=1}^{N} \frac{\mathrm{qtf}_{t,n}}{\log_2(n+1)} \tag{5}$$

ここで，$\mathrm{qtf}_{t,n}$ は索引語 t が N-best リストの n 番目の仮説に出現した回数を表し，N は N-best リストに含まれる仮説数を表す．qtf_t が小数部を持った場合，最も近い整数に切り上げを行う．なお，$N = 1$ の場合には，1-best の仮説の単語数に基づき検索要求を生成した結果に一致する．N-best（対数）に基づいた仮説統合の例を図 1 に示す．図 1 の “A”，“B”，“C”，“D”，“E” はそれぞれ単語を表す．“B” と “E” に着目すると，“B” が初めて出現した仮説は 1 番目の仮説であり，検索要求のベクトルの要素は 2 とな

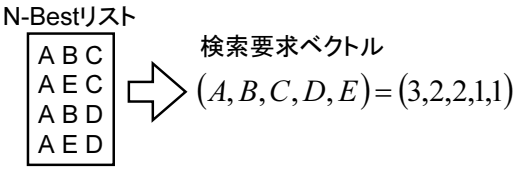

図 1 仮説の順位に基づく重み付きの仮説統合の例（N-best（対数））

る．一方，“E” が初めて出現した仮説は 2 番目の仮説であり，検索要求のベクトルの要素は 1 となる．例に示すとおり，上位の仮説に含まれる索引語の出現数が相対的に高い値となる．

3.3.2 仮説のアライメントに基づく仮説統合

次に, N-best リストをそのまま用いるのではなく, N-best リストから WTN（Word Transition Network）を生成し，各索引語の出現頻度 qtf_t を調整する方法も実現する．具体的には ROVER 法 [10] で用いられている手法で N-best リストを WTN に変換し，その WTN を探索することで qtf_t を求める．WTN の生成手順を以下に示す．

(1) 音声認識の評価関数に基づき仮説を N 個用意（N-best リストを生成）する．各仮説を上位から $W_1, W_2 \ldots W_N$ とする．
(2) W_1 を WTN_1 とする．
(3) $n = 2 \ldots N$ において，DP マッチングを用いて WTN_{n-1} と W_n のアライメントを求め，WTN_n を生成する．
(4) WTN_N の各クラスタ i で，各索引語 t にスコア $S_{i,t}$ を付与する．その際，索引語以外の単語は存在しなかったものとして扱い，そのスコアは NULL 遷移スコア $S_{i,NULL}$ の計算に用いる．
(5) 全クラスタの全索引語にスコアが付与された WTN_N を WTN として出力する．

手順 (4) におけるスコア $S_{i,t}$ は式（6）に基づき計算する．

$$S_{i,t} = \frac{CM_{i,t}^{\gamma_1} \cdot CNT_{i,t}^{\gamma_2}}{\sum_t CM_{i,t}^{\gamma_1} \cdot CNT_{i,t}^{\gamma_2}} \tag{6}$$

ここで，$CM_{i,t}$ はクラスタ i の索引語 t の音声認識時の信頼度を表す．$CNT_{i,t}$ はクラスタ i 内での索引語 t の出現数を表す．γ_1 と γ_2 はそれぞれ $CM_{i,t}$ と $CNT_{i,t}$ の重みパラメータである．

WTN 生成の例を図 2 に示す．図 2 の “A”，“B”，“C”，“D”，“E” はそれぞれ単語を表し，コロンの後の数値は各単語のクラスタ内でのスコアを表す（例えば，“B:0.4” は，単語 “B” のスコアが 0.4 であることを表す）．

次に，WTN からの検索要求の生成手法について述べる．本論文では以下の 3 つを提案する．

- WTN（デコード）：各クラスタの最もスコアの高い単語のみを抽出して qtf_t を与える方法
- WTN（スコア）：各クラスタ内で各索引語に，そのスコアに応じて qtf_t を与える方法
- WTN（枝刈り）：WTN（スコア）と同様にスコアに応じて qtf_t を与えるが，低スコアの索引語には qtf_t を与えない方法

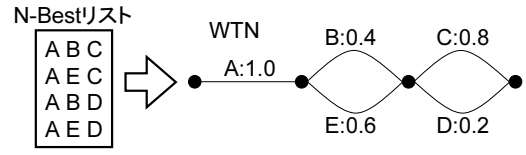

図 2 WTN を用いた仮説統合の例

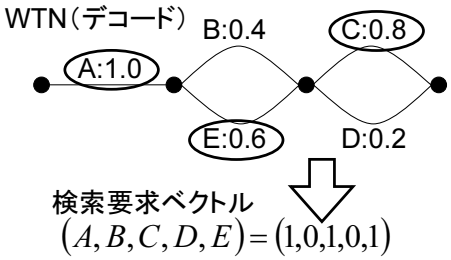

図 3 WTN のデコードに基づく検索要求生成の例

それぞれについて詳細に述べる．

WTN（デコード）

WTN（デコード）は，WTN のクラスタごとに最もスコアが高い単語を選択する手法である．具体的には，索引語 t の出現数 qtf_t を式（7）に基づき計算する．

$$\mathrm{qtf}_t = \sum_{i=1}^{M} \mathrm{IsMax}_{i,t} \tag{7}$$

$$\mathrm{IsMax}_{i,t} = \begin{cases} 1 & \text{if } S_{i,t} = \max_t S_{i,t} \\ 0 & \text{otherwise} \end{cases}$$

ここで，M は WTN のクラスタの総数を表す．$S_{i,t}$ は式（6）で定義されるクラスタ i での索引語 t のスコアである．$\mathrm{IsMax}_{i,t}$ は，$S_{i,t}$ がクラスタ i で最もスコアが高い場合に 1，それ以外の場合は 0 を返す関数である．WTN（デコード）の例を図 3 に示す．この例では，WTN は “A E C” とデコードされ，これに基づき検索要求が生成される．

WTN（スコア）

WTN（スコア）は，WTN のクラスタごとにスコアに応じて索引語の出現数を決定する手法である．具体的には，索引語 t の出現数 qtf_t を式（8）に基づき計算する．

$$\mathrm{qtf}_t = K \sum_{i=1}^{M} S_{i,t} \tag{8}$$

M は WTN のクラスタの総数，$S_{i,t}$ は式（6）で定義されるものである．K は索引語の出現数を整数化するためのパラメータであり，本論文では WTN 生成に使う仮説数 N とする．qtf_t が小数部を持った場合，小数

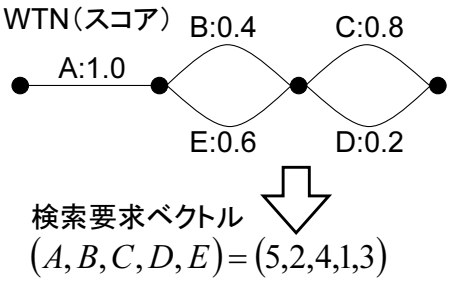

図 4 WTN のスコアに基づく重みを用いた検索要求生成の例（式（8）の $K=5$）

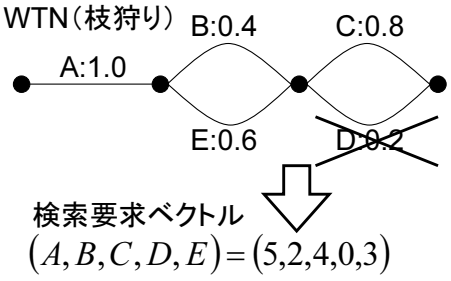

図 5 WTN の枝刈りを用いた検索要求生成の例（式（9）の $K=5$, $\alpha=3$）

第 1 位で四捨五入を行う．WTN（スコア）の例を図 4 に示す．2 番目のクラスタに着目すると，“B” と “E” のスコアはそれぞれ 0.4 と 0.6 である．$K=5$ とすると，“B” の出現数は 2，“E” の出現数は 3 となる．

WTN（枝刈り）

WTN（枝刈り）も WTN（スコア）と同様に WTN のクラスタごとにスコアに応じて索引語の出現数を決定する手法である．ただし，各クラスタで低スコアの索引語の出現数を 0 とするものである．具体的には，索引語 t の出現数 qtf_t を式（9）に基づき計算する．

$$\mathrm{qtf}_t = K \sum_{i=1}^{M} \mathrm{Score}_{i,t} \tag{9}$$

$$\mathrm{Score}_{i,t} = \begin{cases} S_{i,t} & \text{if } \dfrac{\max\limits_t S_{i,t}}{S_{i,t}} \leq \alpha \\ 0 & \text{otherwise} \end{cases}$$

M は WTN のクラスタの総数，$S_{i,t}$ は式（6）で定義されるものである．K は索引語の出現数を整数化するためのパラメータであり，本論文では WTN 生成に使う仮説数 N とする．$\mathrm{Score}_{i,t}$ は $S_{i,t}$ とクラスタ i 内で最も高いスコアとの比がしきい値 α 以下のときは $S_{i,t}$，それ以外のときは 0 を返す関数である．qtf_t が小数部を持った場合，小数第 1 位で四捨五入を行う．なお，しきい値 α を無限大とすると，式（9）は式（8）に一致する．WTN（枝刈り）の例を図 5 に示す．図 5 の 3 番目のクラスタに着目すると，このクラスタ内で最も高いスコアを持つ索引語は “C” であり，しきい値 $\alpha=3$ すると，“C” と “D” のスコアの比は 4 ($>\alpha$) であるため “D” は検索要求に含まれない．

4. システムの評価尺度

4.1 音声認識の評価尺度

音声認識の評価尺度として，WER および WWER（式（1））を用いた．WWER 計算時の単語重要度として，教師なし推定 [11] により推定した重要度を与えた．この推定手法は，検索性能に大きな影響を与える単語に大きな重要度を与えるものである．

4.2 情報検索の評価尺度

4.2.1 11 点平均精度

情報検索の評価尺度として，11 点平均精度（11ptAP: 11-point Average Precision）[12] を用いた（式（10））．

$$\mathrm{11ptAP}_{Q_k} = \frac{1}{11} \sum_{i=0}^{10} IP_{Q_k}\left(\frac{i}{10}\right) \tag{10}$$

$$IP_{Q_k}(x) = \max_{x \leq R_{Q_k}(t)} P_{Q_k}(t)$$

ここで，$R_{Q_k}(t)$ と $P_{Q_k}(t)$ は，それぞれ Q_k に関する検索順位 t における再現率と精度を表す．$IP_{Q_k}(x)$ は，再現率レベルが x 以上の精度 $P_{Q_k}(t)$ の最大値を表す補間精度である．

4.2.2 音声認識誤りによる検索性能低下率

前項で述べたような情報検索の評価スコアは，情報検索システム自体の性能の影響を受け，たとえ音声認識誤りが 0 であっても，検索性能は最高の値（11 点平均精度であれば 1）にならない．そこで本論文では，音声認識による検索性能の低下を評価するための尺度として検索性能低下率（IRDR: Information Retrieval performance Degradation Ratio）[11] を用いる（式（11））．

$$\mathrm{IRDR} = 1 - \frac{H}{R} \tag{11}$$

R と H はそれぞれ，書き起こしと音声認識結果の検索要求を用いた際の検索性能（本論文では 11 点平均精度）を表す．音声認識誤りがない場合は IRDR は 0 となり，IRDR は音声認識誤りによる情報検索の性能低下の割合を表す尺度となる．

5. 評価実験

5.1 音声検索システム

音声入力型情報検索システムを構築し評価を行った．本節では構築した音声検索システムについて述べる．

5.1.1 音声認識システム

音声認識システムのデコーダには，Julius rev.4.1.5.1 に MBR 機能を実装した MBR-Julius[13] を用いた．音響モデルには，JNAS コーパスから学習した triphone モデル（CSRC2003 年度最終版 [14] に収録）を用いた．言語モデルには，CSJ[15] の講演 2702 件の書き起こしから学習した

3-gram 言語モデル（語彙サイズ約 20K）を用いた．

音声認識手法として，以下の 3 種類を用いた．

- 事後確率最大化音声認識（ベースライン）
- 単語の誤り数を最小化する WER 最小化音声認識
- 検索の重要語の誤りを最小化する WWER 最小化音声認識

WER 最小化および WWER 最小化音声認識は，ベースライン音声認識により 100-best 仮説を生成し MBR 基準でリスコアリングして行った．MBR 音声認識のためのゆう度と損失関数に関するパラメータは，当該データを用いた音声認識実験において MBR 音声認識の精度が高くなったものとした．その他の音声認識パラメータには Julius rev.4.1.5.1 のデフォルト値をそのまま用いた．

5.1.2 検索要求の生成手法

ベースライン手法として 1-best 仮説から検索要求を生成する方法を用いた．複数仮説から検索要求を生成する手法として，N-best（一様），N-best（線形），N-best（対数），WTN（デコード），WTN（スコア），WTN（枝刈り）を実装した．全てにおいて用いる仮説数 N は 5 とした．また，WTN 生成のためのパラメータ，式（6）の γ_1 と γ_2，および式（9）の α は，開発セット，具体的には交差検定（leave-one-out）によって決定した．

5.1.3 情報検索システム

情報検索システムとしてベクトル空間モデルに基づく文書検索システムを採用し，GETA[16] を用いて構築した．索引語には名詞と動詞の基本形を用いた．本研究では，検索要求 $\boldsymbol{Q}$ が与えられたとき，全ての文書 $\boldsymbol{D}_i$ について $\boldsymbol{Q}$ との類似度 $\mathrm{Sim}(\boldsymbol{Q}, \boldsymbol{D}_i)$ を算出し，類似度が高い順に上位 1000 件を出力することとした．

本研究では，ベクトルの類似度尺度として SMART[17] を用いた（式（12））．

$$\mathrm{Sim}\left(\boldsymbol{Q}, \boldsymbol{D}_i\right) = \mathrm{SMART}\left(\boldsymbol{Q}, \boldsymbol{D}_i\right) = \sum_{t=1}^{T} \left(Q_t \cdot D_{i,t}\right) \quad (12)$$

$$Q_t = \begin{cases} \dfrac{1+\log\left(\mathrm{qtf}_t\right)}{1+\log\left(\mathrm{avqtf}\right)} \cdot \log\dfrac{N_{doc}}{n_t} & \text{if } \mathrm{qtf}_t > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$D_{i,t} = \begin{cases} \dfrac{1+\log\left(\mathrm{tf}_{i,t}\right)}{1+\log\left(\mathrm{avtf}\right)} \cdot \mathrm{Norm} & \text{if } \mathrm{tf}_{i,t} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{Norm} = \frac{1}{(1-\mathrm{slope}) \cdot \mathrm{pivot} + \mathrm{slope} \cdot \mathrm{utf}_i}$$

ここで，$\mathrm{tf}_{i,t}$ は $\boldsymbol{D}_i$ 中での索引語 t の出現数，avtf は $\boldsymbol{D}_i$ における索引語の出現数の平均を表す．pivot は 1 文書中の異なり索引語数の平均，utf_i は $\boldsymbol{D}_i$ 中の異なり索引語数を表す．slope は補間係数であり，本研究では 0.2 とした．qtf_t は，$\boldsymbol{Q}$ 中での索引語 t の出現数，avqtf は $\boldsymbol{Q}$ に含まれる索引語の出現数の平均を表す．N_{doc} は検索対象の文書集合の全文書数を表し，n_t は索引語 t を含む文書の数を表す．T は索引語の総数を表す．

通常の N-Best	1. 道路 / 等 / の / 発生 / する / 助言 / を / 知り / たい 2. 道路 / 等 / の / 発生 / する / 条件 / が / 知り / たい 3. オーロラ / の / 派生 / する / 条件 / を / し / たい 4. オーロラ / の / 反省 / する / 上限 / が / し / たい 5. **オーロラ / の / 発生 / する / 条件 / が / 知り / たい**
WWER 最小化後の N-Best	1. オーロラ / の / 派生 / する / 条件 / を / し / たい 2. 道路 / 等 / の / 発生 / する / 助言 / を / 知り / たい 3. **オーロラ / の / 発生 / する / 条件 / が / 知り / たい** 4.道路 / 等 / の / 発生 / する / 条件 / が / 知り / たい 5. オーロラ / の / 反省 / する / 上限 / が / し / たい

図 6 事後確率最大化音声認識と WWER 最小化音声認識の N-best の比較

5.2 検索タスク

検索タスクとして，日本語音声ドキュメント検索テストコレクション [18] を用いた．これは，情報処理学会音声言語情報処理研究会の音声ドキュメント処理ワーキンググループが作成した音声ドキュメント検索評価用テストコレクションである．テストコレクションには研究者が共通で使える実験データセット（講演音声データ，検索課題とそれに対する正解データ）が用意されている．データセットの構成を以下に示す．

- 検索対象文書 : CSJ の講演の書き起こしテキスト（2702 講演）
- 利用者の検索要求を記述した「検索課題」: 39 課題
- 検索課題を満たす「正解文書のリスト」

テストコレクションには検索課題の読み上げ音声データが含まれていないため，検索課題の音声データとして，男性 10 名と女性 4 名の計 14 名に読み上げてもらった合計 546 件 [19] を用いた．

5.3 実験結果

5.3.1 N-best リストを用いる効果

事後確率最大化音声認識および WWER 最小化音声認識を行った際の N-best リストの比較を図 6 に示す．なおこれは，説明のために作成した例であり，実例ではない．この例では，通常の音声認識結果の 5 番目に適切と考えられる仮説（図中の下線を付与している仮説）が出現している．一方，WWER 最小化音声認識の結果には，適切な仮説は 3 番目に出力されている．このように，WWER 最小化音声認識を行えば，検索の観点で適切な仮説が 1-best に出現しない場合であっても，上位の仮説として出現しやすくなると考えられる．

実際に，事後確率最大化音声認識および MBR 音声認識（WWER 最小化音声認識）を行い，これにより生成した N-best リストの評価を行った．具体的には，各音声認識の結果得られた N-best リストの 1-best の仮説の認識率（WWER）の平均，2-best の仮説の認識率の平均，...，100-best の仮説の認識率の平均を求めた．その結果を図

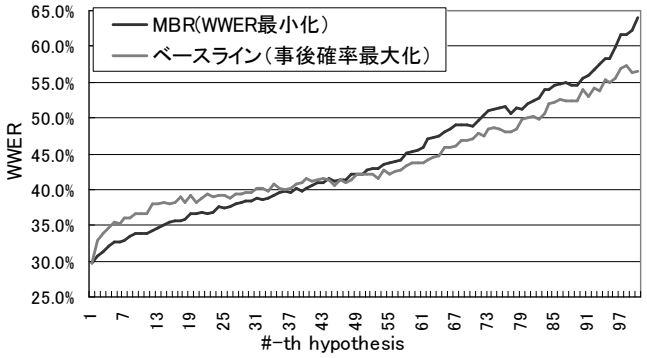

図 7 事後確率最大化音声認識と WWER 最小化音声認識の N-best の音声認識率の比較

7 に示す．1-best での認識率は同等であるものの，2-best 以降の結果では，上位の仮説の認識率は MBR 音声認識を行った場合に認識率が高くなる（WWER が低くなる）傾向が見られている．このことから，MBR 音声認識を行うことで N-best リスト中の上位に認識率の高い仮説を得られやすくなる，すなわち上位の質の高い N-best リストを生成できることがわかる．このことは，MBR 音声認識（WWER 最小化）を行って N-best リストを生成したうえで上位の仮説を用いて検索要求を生成することで，より適切な検索要求ベクトルを生成できる可能性を示している．

5.3.2 実験結果

はじめに，1-best 仮説のみから検索要求を生成して検索を行った結果を表 1 に示す．書き起こしテキスト（音声認識誤りなしに相当）を用いて検索を行った場合の検索精度は 0.428 であった．従来の音声認識（事後確率最大化音声認識）の結果を用いて検索を行った場合は，検索精度は 0.358 であり，IRDR は 16.4%であった．MBR 音声認識で WER 最小化を行った場合は，WER は 26.6%から 26.4%と改善が得られた．MBR 音声認識で WWER 最小化を行った場合は，WWER は 31.6%から 31.3%と改善が得られた．MBR 音声認識により音声認識精度は改善されているものの，検索精度は 0.358 （IRDR 16.4%）と変わらなかった．このことは，実際に 1-best 仮説の音声認識誤りは改善できるものの，それだけを用いても検索精度の向上につながらないタスクがあることを示している．

次に N-best リストから検索要求を生成して検索を行った結果を表 2 に示す．事後確率最大化音声認識の N-best リストを用いて検索要求を生成した場合には，いずれの手法でも 1-best 仮説のみを用いたときよりも検索精度が低下（IRDR が増加）した．これは，MBR 音声認識を行わずに N-best リストを用いて検索要求を生成しても，検索性能の向上を得ることが難しいことを示している．MBR 音声認識（WER 最小化）の N-best リストを用いて検索要求を生成した場合は，WTN（デコード）の生成手法を用いたときにのみ IRDR の改善が得られ，IRDR 15.9%が得られた．重要度を用いた MBR 音声認識（WWER 最小化）の結果

表 1 1-best 仮説のみからの検索要求生成の結果

	音声認識手法			テキスト
	事後確率	MBR		（音声認識
	最大化	WER	WWER	誤りなし）
11ptAP	0.358	0.358	0.358	0.428
IRDR(%)	16.4	16.4	16.4	0

表 2 N-best リストからの検索要求生成の結果（IRDR(%)）

検索要求	音声認識手法		
	事後確率	MBR	
生成手法	最大化	WER	WWER
N-best（一様）	18.0	17.8	16.8
N-best（線形）	17.3	17.3	16.8
N-best（対数）	17.5	17.1	16.6
WTN（デコード）	19.9	**15.9**	**16.1**
WTN（スコア）	17.8	16.8	**15.9**
WTN（枝刈り）	18.9	16.6	**15.4**

太字は，1-best のみを用いたときよりも IRDR が低い（検索性能の向上が得られている）ものを示す．

から検索要求を生成した場合は，WTN（デコード），WTN（スコア），WTN（枝刈り）の生成手法を用いたときに検索性能の改善が得られ，IRDR はそれぞれ 16.1%，15.9%，15.4%となった．MBR 音声認識（WER や WWER 最小化）を行った上で WTN を構築し，検索要求を生成することにより，検索性能の低下率を抑えることができた．特に，MBR 音声認識（WWER 最小化）と WTN（枝刈り）の組み合わせによって，本研究で最大の IRDR の改善（約 6%: 16.4%→15.4%）が得られた．

WTN からの検索要求の生成は，N-best リスト中に含まれている索引語のうち，信頼度の低いものを検索要求に含めない，もしくは検索要求中での出現頻度を小さくする手法であり，このような検索要求生成法が有効であることを示している．

次に，各音声認識手法について比較を行う．N-best（対数）に着目すると，事後確率最大化音声認識，WER 最小化音声認識，WWER 最小化音声認識の結果（IRDR）はそれぞれ 17.5%，17.1%，16.6%であり，WWER 最小化音声認識を用いた際に最も IRDR が低く，検索性能が高い．N-best（一様），N-best（線形），WTN（デコード），WTN（スコア），WTN（枝刈り）においても，事後確率最大化音声認識よりも WER 最小化音声認識，WWER 最小化音声認識の結果で IRDR が低い．また，WWER 最小化音声認識を行ったときは WER 最小化音声認識を行ったときよりも IRDR が低い，もしくは同等である．この結果は，WWER 最小化音声認識を行うことで N-best リストの上位候補の質が向上し，適切な検索要求を生成することができるようになることを示している．

WWER 最小化音声認識を行って質の高い N-best リストを生成してから WTN を構築した後に，WTN から信頼

表 3 提案法（WWER 最小化+WTN（枝刈り））とベースライン（事後確率最大化+1-best）の比較

検索精度	検索要求数	11ptAP	WWER
向上	175	0.266→0.305	42.3%→41.2%
低下	112	0.270→0.228	45.8%→47.3%
変化なし	259	0.458→0.458	23.7%→22.4%

度を考慮して検索要求を生成する効果を示した．

5.3.3 実験結果の分析

提案法（WWER 最小化+WTN（枝刈り））を行ったときと，ベースライン（事後確率最大化+1-best 利用）との結果を比較した．

546 件の検索要求のうち，提案法により検索精度が向上/低下した検索要求を調べた．結果を表 3 に示す．175 件の検索要求で検索精度が向上し（11ptAP: 0.266→0.305），112 件の検索要求で検索精度が低下した（11ptAP: 0.270→0.228）．残りの 259 件は検索精度に変化がなかった（11ptAP: 0.458→0.458）．検索精度が低い時に提案法により変化が得られていることがわかる．

次に，これらのグループごとに上位の 5-best 仮説の音声認識率（WWER）も調べた．ここでは，各検索要求について 1-best 仮説から 5-best 仮説までのそれぞれの WWER を求めてその平均をとったものを各検索要求に対する WWER とし，それらの平均を求めた．結果は表 3 に示されている．検索精度に変化がなかったグループ（259 件）では，音声認識率（WWER）はベースラインの音声認識 23.7%に対して MBR 音声認識（WWER 最小化）では 22.4%であった．検索精度に変化があったグループでは，検索精度が向上したグループ（175 件）での WWER はベースライン認識で 42.3%，MBR 音声認識（WWER 最小化）で 41.2%であり，検索精度が低下したグループ（112 件）での WWER はベースライン認識で 45.8%，MBR 音声認識（WWER 最小化）で 47.3%であった．検索結果に変化があったときは，もともとの音声認識精度が低かった（WWER が高かった）ことがわかる．MBR 音声認識は，音声認識率が低いときに効果があることが知られており [20]，この結果もそれに一致する．

最後に，提案法とベースライン法の間で検索精度に差があるかについて符号検定（有意水準 1%）を行った．提案法とベースライン法の間に有意な差がみられ，提案法に効果があることがわかった．

もともと音声認識精度が低く検索精度が低いような場合に，たとえば認識精度が低い話者などに対して，本提案手法は効果的と考えられる．

6. おわりに

音声入力型情報検索のための音声認識手法と検索要求生成手法について検討を行った．具体的には，ベイズリスク最小化音声認識を行い，その結果得られた N-best リストを用いて検索要求を生成する手法を提案した．実験の結果，WWER 最小化音声認識を行って重要単語の誤りが少ない N-best リストを生成し，その N-best リストの上位を用いて WTN を構築して信頼度を考慮することで，適切な検索要求が生成できることを示した．

謝辞 本研究は科研費の助成を受けた．

参考文献

[1] 翠 輝久，河原達也：限定されたドメインにおける質問応答機能を備えた文書検索・提示型対話システム，情報処理学会研究報告，2006-SLP-62, pp. 69–74 (2006).
[2] 桐山伸也，広瀬啓吉，峯松信明：話題知識を導入した文献検索音声対話システム，電子情報通信学会論文誌，Vol. J85-D-II, No. 5, pp. 863–876 (2002).
[3] Matsushita, M., Nishizaki, H., Utsuro, T. and Nakagawa, S.: Improving Keyword Recognition of Spoken Queries by Combining Multiple Speech Recognizer's Output for Speech-driven WEB Retrieval, *IEICE TRANS. & SYST.*, Vol. E88-D, No. 3, pp. 472–480 (2005).
[4] 南條浩輝，河原達也，七里 崇：音声理解を指向したベイズリスク最小化枠組みに基づく音声認識，電子情報通信学会論文誌，Vol. J91-D, No. 5, pp. 1314–1324 (2008).
[5] 松尾宏規，西田昌史，古谷 遼，南條浩輝，山本誠一：単語の重要度を考慮したベイズリスク最小化音声認識を用いた音声入力型情報検索システムの評価，日本音響学会講演論文集，秋季研究発表会，pp. 201–202 (2011).
[6] Goel, V., Byrne, W. and Khudanpur, S.: LVCSR rescoring with modified loss functions: A decision theoretic perspective, *Proc. IEEE-ICASSP*, Vol. 1, pp. 425–428 (1998).
[7] Stolcke, A., König, Y. and Weintraub, M.: Explicit word error minimization in N-best list rescoring, *Proc. EUROSPEECH*, pp. 163–166 (1997).
[8] 松下雅彦，西崎博光，宇津呂武仁，中川聖一：音声入力による Web 検索のためのキーワード認識・抽出法の検討，情報処理学会研究報告，2003-SLP-48, pp. 21–28 (2003).
[9] 西崎博光，中川聖一：音声キーワードによるニュース音声データベース検索手法，情報処理学会論文誌，Vol. 42, No. 12, pp. 3173–3184 (2001).
[10] Fiscus, J.: A post-processing system to yield reduced word error rates: Recognizer Output Voting Error Reduction (ROVER), *Proc. IEEE-ASRU*, pp. 347–354 (1997).
[11] 古谷 遼，七里 崇，南條浩輝：音声入力型情報検索におけるベイズリスク最小化音声認識のための単語重要度の自動推定，情報処理学会論文誌（採録決定），Vol. 54, No. 7 (2013).
[12] 北 研二，津田和彦，獅々堀正幹：情報検索アルゴリズム，共立出版.ISBN 4-320-12036-1.
[13] 南條浩輝，古谷 遼，西田昌史：オープンソース音声認識エンジン Julius へのベイズリスク最小化機能の実装と評価，電子情報通信学会論文誌（採録決定），Vol. J96-D, No. 10 (2013).
[14] 河原達也，武田一哉，伊藤克亘，李 晃伸，鹿野清宏，山田 篤：連続音声認識コンソーシアムの活動報告及び最終版ソフトウェアの概要，情報処理学会研究報告，2003-SLP-49, pp. 325–330 (2003).
[15] Maekawa, K.: Corpus of Spontaneous Japanese: Its design and evaluation, *Proc. ISCA & IEEE-SSPR*, pp. 7–12 (2003).
[16] 高野明彦，西岡真吾，今一 修，岩山 真，丹羽芳樹，久光 徹，藤尾正和，徳永健伸，奥村 学，望月 源，野本

忠司：汎用連想計算エンジンの開発と大規模文書分析への応用 (2002).http://geta.ex.nii.ac.jp/pdf/itx002.pdf.

[17] 小作浩美，内山将夫，井佐原均，河野恭之，木戸出正継：WWW 検索における複数検索結果の統合処理とその評価，情報処理学会論文誌，Vol. 44, No. SIG 8(TOD 18), pp. 78–91 (2003).

[18] Akiba, T., Aikawa, K., Itoh, Y., Kawahara, T., Nanjo, H., Nishizaki, H., Yasuda, N., Yamashita, Y. and Itou, K.: Construction of a Test Collection for Spoken Document Retrieval from Lecture Audio Data, *IPSJ-journal*, Vol. 50, No. 2, pp. 82–94 (2009).

[19] 七里　崇，重安幸治，南條浩輝，吉見毅彦：音声クエリによる講演音声ドキュメント検索の基礎的評価，第 4 回音声ドキュメント処理ワークショップ，No. 16 (2010).

[20] Schlüter, R., Nussbaum-Thom, M. and Ney, H.: On the relation of Bayes Risk, Word Error, and Word Posteriors in ASR, *Proc. INTERSPEECH*, pp. 230–233 (2010).