Vol.2011-NL-201 No.5
Vol.2011-SLP-86 No.5
2011/5/16

音声入力型大学情報検索システムに対するベイズリスク最小化音声認識の適用

松尾宏規† 西田昌史† 古谷遼†† 南條浩輝†† 山本誠一†

音声入力型の情報検索では検索クエリ中の重要語句を正確に認識する必要があり，それらの認識誤りを少なくすることが重要である．しかし，従来の音声情報検索においては尤度最大化音声認識が用いられており，単語誤りについて考慮されていない．そこで本研究では，単語誤り率の最小化を行うベイズリスク最小化音声認識を音声入力による大学情報検索システムに導入した．本手法の有効性を示すために評価実験を行った結果，従来の尤度最大化音声認識に比べてベイズリスク最小化音声認識により音声認識精度ならびに検索精度を改善することができた．

Collage Information Retrieval System based on Minimum Bayes-Risk Decording

HIROKI MATSUO† MASAFUMI NISHIDA†
RYO FURUTANI†† HIROAKI NANJO††
SEIICHI YAMAMOTO†

In information retrieval based on spoken queries, it is important to recognize words in the spoken queries correctly. However, the conventional information retrievals based on spoken queries have not taken recognition errors into account because it has used the speech recognition based on maximum likelihood estimation. We propose a collage information retrieval system based on minimum Bayes-risk decording which minimizes the word error rate. To evaluate effectiveness of the proposed method, we conducted experiments. From experimental results, we demonstrated that the proposed method can improve the speech recognition accuracy and information retrieval accuracy compared with the conventional speech recognition based on maximum likelihood estimation.

1.はじめに

インターネット上のデータ量が増大し，情報検索サービスはインターネットにおいて必要不可欠な技術になった．テキストデータの検索に関しては多数のウェブサービスが公開され，キーワードをタイプするだけで全文検索をすることができる．そして近年，話し言葉の進展に伴い[1][2]，音声認識技術が情報検索に応用される事が多くなった．Web ページの情報を調べる際にサイト内での検索が困難な場合，必要な情報を得るためにはホームページをツリー上にたどっていかなければならないが，音声入力による情報検索はショートカットすることができるという点で効率的だと考えられる．

ニュースや講演音声などの音声メディアを検索対象とし入力がテキストのものとしては，ニュース音声記事の検索に単語ベースと音節ベースのインデクスを用いる研究[3]，音声認識による自動書き起こしと人手書き起こしの差異を翻訳によって補完する検索手法の提案に関する研究[4]，サブワードを用いた語彙制限のない音声文書検索システムを目指す研究[5]，WEB テキストを使用してインデクス拡張を行い検索精度の改善を目指す研究[6]等が行われている．また，入力が音声であるものとしては，音声文書を対象とした音声情報検索システム[7]や講演音声のドキュメント検索に関する研究[8]が存在する．

本研究では音声入力型で Web ページを対象とした情報検索システムの構築を目指している．音声入力型の情報検索ではクエリ中の重要語句を正確に認識する必要があり，それらの認識誤りを少なくすることが重要である．しかし，従来の音声情報検索においては尤度最大化音声認識が用いられており．尤度を基準として認識しているため単語誤りについては考慮されていない．そこで本研究では，単語誤り率の最小化を行うベイズリスク最小化音声認識[9]を導入した音声入力型情報検索について検討を行う．

本論文では 2 章で今回情報検索の対象として使用した同志社大学のホームページの構造，そして構築した音声情報検索システムの概要を述べる．3 章では導入したベイズリスク最小化音声認識について述べる．4 章では検索手法として採用したベクトル空間法，検索質問と各ホームページの類似度として採用した cosine 尺度について述べる．5 章で実際に大学情報検索システムにベイズリスク最小化音声認識を導入して実験した結果と考察について述べ，6 章でまとめについて述べる．

† 同志社大学
Doshisha University
†† 龍谷大学
Ryukoku University

Vol.2011-NL-201 No.5
Vol.2011-SLP-86 No.5
2011/5/16

2.ベイズリスク最小化音声認識に基づく大学情報検索システム

今回はベイズリスク最小化音声認識を音声情報検索システムに導入するにあたり，同志社大学の WEB ページ検索システムを構築した．図 1 に同志社大学のホームページの構造を示す．同志社大学のホームページは大きく分けると大学紹介，教育，研究活動，留学に分かれている．教育は主に学部学科，大学院の研究科と専攻に関する紹介のページになっている．一般的にホームページは木構造になっており，情報の粒度は下位の階層に行くにつれ細かくなる．音声認識を使用しない場合，仮にユーザが哲学科のホームページを閲覧したいと思った際はトップページ→教育→文学部→哲学科とリンクを探してたどっていく必要があるが、音声認識を使用すると「文学部哲学科について知りたい」などと発話すると哲学科のページに飛ぶことができ，リンクを探す手間を省くことができる．

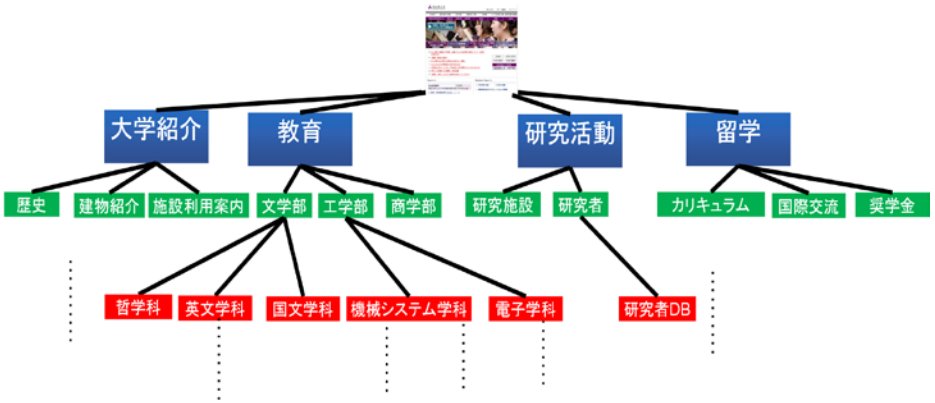

図 1　同志社大学のホームページの構造

図 2 に大学情報検索システムの概要を示す．まずユーザの音声クエリを尤度最大化音声認識で認識する．そして求めた N-best リストに対しベイズリスク最小化音声認識を施し、認識スコアを再計算する．ベイズリスク最小化音声認識で再計算された認識スコアの最も高い認識結果を検索クエリとして検索を行う．情報検索には従来からよく使用されているベクトル空間法を使用し、事前に大学のホームページから単語を抽出し，その頻度を要素としたベクトルを用意しておく．そして入力されたクエリの認識結果から単語を抽出しベクトル化する．これらのベクトルの類似度が高いページを検索結果として出力する．

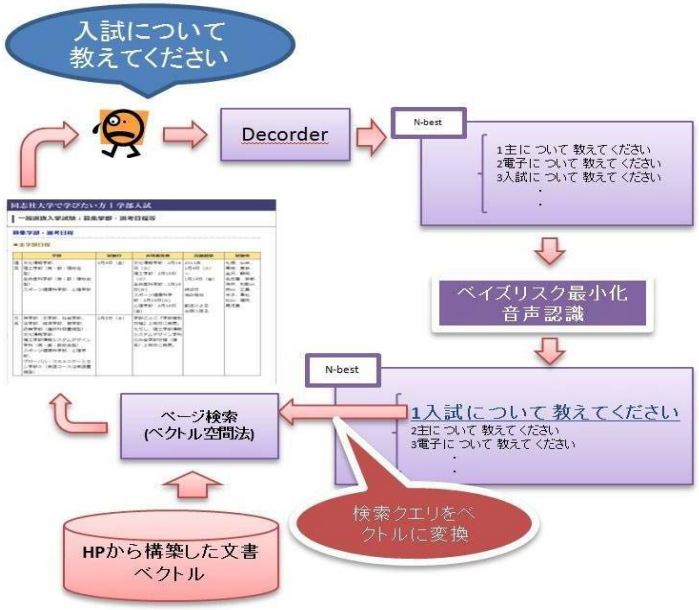

図 2　音声入力型大学情報検索システムの概要

Vol.2011-NL-201 No.5
Vol.2011-SLP-86 No.5
2011/5/16

3.ベイズリスク最小化音声認識

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(1)に表す．W'は認識結果の候補になる文字列，$\hat{W}$は音声認識により認識結果として出力される単語列である．

$$\hat{W} = \underset{W'}{\operatorname{argmax}}\, P(W'|X) \qquad (1)$$

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].

$$\delta(X) = \underset{W}{\operatorname{argmin}} \sum_{W'} l(W,W') \cdot P(W'|X) \qquad (2)$$

さらに，それぞれのスコアに重みパラメータを乗じる手法の有効性が先行研究で示されており[9]，本研究でも重みパラメータを用いる．式(3)に重みパラメータを乗じたものを示す．λ_1は損失関数に対する重み，λ_2は認識スコアに対する重みである．

$$\delta(X) = \underset{W}{\operatorname{argmin}} \sum_{W'} l(W,W')^{\lambda_1} \cdot P(W',X)^{\lambda_2} \qquad (3)$$

WER の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（以下 WER），もしくは,WER の定義式の分子に相当する編集距離(=Lavenshtein Distance)を用いればよいことが知られている[11][12].編集距離とは文字の挿入や削除，置換によって一つの文字列を別の文字列に変形するのに必要な手順の最小回数として与えられる．これは言い換えると二つの文字列がどの程度異なっているかを示す数値である． WER は式(4)で定義される．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 は削除誤りの単語の数、I は挿入誤りの単語の数である．

$$\mathrm{WER} = \frac{I + D + S}{N} \qquad (4)$$

3.2 N-best リストに基づく音声認識

本研究で用いた MBR 音声認識のアルゴリズムについて以下に述べる，

(1) 音声クエリに対して尤度最大化音声認識を行い、認識スコアの高い順に仮説を N 個求めて N-best リストを作成する．

(2) そして各仮説の評価値$^{(W_i(i=1...N))}$を式(1)で再計算する

(3) 最も評価値の低いものを認識結果として出力する．評価値の計算は式(5) を使用する．λ1,λ2 はそれぞれ損失関数，認識スコアに対する重みである．

$$f(W_i) = \sum_{W_j \in \text{N-best list}} l(W_i,W_j)^{\lambda_1} \cdot P(W_j)^{\lambda_2} \qquad (5)$$

損失関数には最小編集距離を使用した．MBR 音声認識は音声入力された発話に対して行う．

4.検索手法

検索手法にはベクトル空間法を使用する．ベクトル空間法とは文書を多次元空間上のベクトルとして表現し，二つのベクトルを比較することにより類似度を調べるものである．適用例としては 1960 年から実験されてきた SMART[13]が有名である．本研究では二つのベクトルを文書ベクトルと検索質問ベクトルとした．ホームページテキストを文書ベクトルに変換する際，ホームページのテキストを索引語に分解し，索引語毎の出現頻度を求め，要素とした．出現頻度は検索対象の各ホームページにおける出現数をすべて足したものである．検索クエリを 検索質問ベクトルに変換する際は，MBR 音声認識もしくは尤度最大化音声認識で認識結果として出力した単語列を索引語毎に出現頻度を要素として構築した．文書ベクトルはホームページ 1 ページ毎に作成する．

ホームページテキストを索引語に分解する過程には形態素解析ツールの Chasen[14]を

Vol.2011-NL-201 No.5
Vol.2011-SLP-86 No.5
2011/5/16

使用した．尚，文書と検索質問の類似度は文章の類似度ではなく内容の類似を求めるので，内容の類似度には関係ないと思われる機能語は索引語から取り除いた．

各文書ベクトルと検索質問ベクトルの類似度計算には式(6)のcosine尺度を使用した．この値が高い順にページを出力する．式(7)に文書ベクトルと検索質問ベクトルを具体的に示す．dは文書ベクトル，qは検索質問ベクトル，mは索引語の語彙数を表す． d_{mn}は索引語の頻度，q_mは索引語の頻度，mは索引語の語彙数，nは検索対象文書数を表す．

図 3，図 4 に同志社大学のホームページで出現頻度が多い上位 20 位の索引語を示す．索引語のカウントに使用した対象のホームページは今回評価に使用する同志社大学のホームページ 938 ページである．出現頻度は 938 ページすべてのページにおける出現数を足したものである．索引語の抽出には Chasen を用い，機能語は取り除いてある．

$$\cos(d_j, q) = \frac{d_j \cdot q}{\|d_j\|\|q\|} = \frac{\sum_{i=1}^{m} d_{ij}^2 q_i}{\sqrt{\sum_{i=1}^{m} d_{ij}^2}\sqrt{\sum_{i=1}^{m} q_i^2}} \quad (6)$$

$$D = [d_1 d_2 \ldots d_n] = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix} \qquad q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{bmatrix} \quad (7)$$

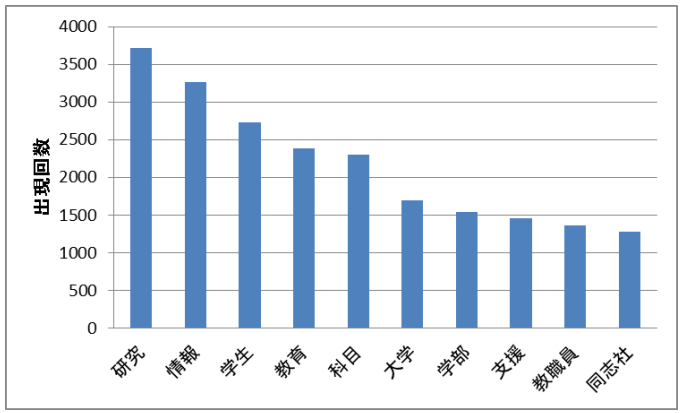

図 3　単語の出現頻度（上位 1~10 位）

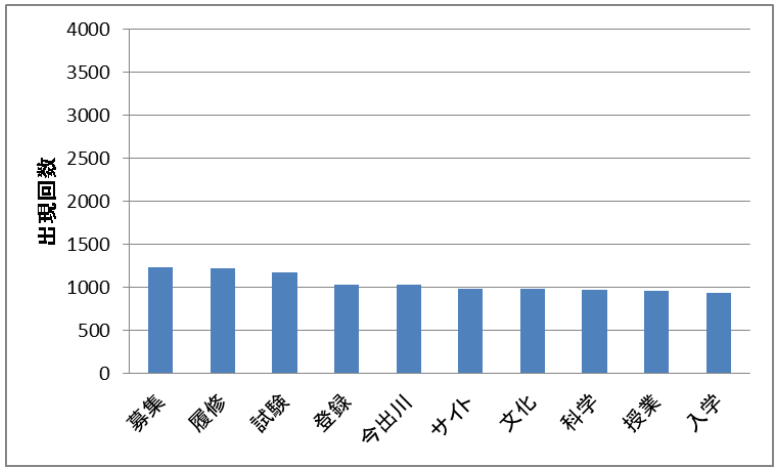

図 4　単語の出現頻度（上位 11~20 位）

5.評価実験

5.1 実験条件

これまでに述べた大学情報検索システムを実装し，評価実験を行った．デコーダは Julius-4.1.5，音響モデルは日本語話し言葉コーパスの 2496 講演から学習された状態共有モデル PTM を使用した．音声分析は 16kHz サンプリング，フレーム長 25msec ハミング窓，フレーム幅 10msec，特徴量は MFCC(12 次)+ΔMFCC(12 次)+ΔPower の計 25 次元である．

言語モデルには同志社大学のホームページ 938 ページと,想定質問文 2396 文から学習した前向き trigram を用いた.語彙数は 15514 である．想定質問文の作成方法は，まず同志社大学のホームページを形態素解析にかけ，名詞のみ抽出し，手動でカテゴリを分ける．カテゴリは研究，進路、同志社大学の施設，同志社大学に関連の深い地域，国名である．そしてカテゴリごとに言い回しを主観的に決め作成した．評価データは 8 名が 14 文を 3 回ずつ発話した計 336 発話である．図 4 に発話した文の例を示す．検索精度に関しては「出力した 10 文書のうちに検索質問の答えに適合したページが存在するとき」を検索成功とした．

Vol.2011-NL-201 No.5
Vol.2011-SLP-86 No.5
2011/5/16

パソコンが使える場所　　今出川キャンパスの行き方を教えて
知真館はどこにありますか　理工学部の学費
水曜の休講を教えて　　今出川キャンパスの行き方を教えて
京田辺のキャンパスマップ　食堂の場所
書籍部の場所　　大学院修士課程の学費
生協のホームページ　　フランス留学の方法
春学期科目の登録日程　　学部入試について教えて

図 4　検索質問として使用した評価データ

5.2 実験結果

表 2　実験結果 1

	従来法	MBR 音声認識
認識精度(%)	91.6	92.4
検索精度(%)	87.5	87.5

表 2 に 5.1 で示した実験条件での実験結果を示す。表中の従来法は，尤度最大化音声認識による実験結果であり，MBR 音声認識はベイズリスク最小化音声認識による実験結果を示している．また，認識精度は単語認識精度を示している． ベイズリスク最小化音声認識では，式(5)における認識スコアと損失関数の重み付けはそれぞれ 1.0,0.1 であり，N-best は 300 ベストと設定した．検索クエリを書き起こしテキストで行ったもの（誤認識を全く含んでいないクエリ）での検索精度は 100%だった．

ベイズリスク最小化音声認識により認識精度に若干の向上は見られたが，検索精度の向上は見られなかった．これは検索クエリ中の内容語の認識精度は向上せず，機能語(その単語のみでは意味を成さない物、助詞等)の認識精度のみが向上されたのが原因であると考えられる．

更に，検索対象のホームページを増やし実験を行った．ページ数は 3772 ページ，語彙数は 25928 となった．5.1 で使用したホームページの 938 ページはすべてこの中に含まれている．使用した音響モデルと音声クエリは 5.1 と同じものである．使用した言語モデルは，大学情報検索の対象にしたホームページ 3772 ページ+想定質問文 2396 文を使って再構築を行った。語彙数は 25949 であった．

表 3 に実験結果を示す．語彙を増やすと語彙が少ない時と比べて認識精度は低下したが，ベイズリスク最小化音声認識により音声認識精度ならびに検索精度を改善することができた．具体的に内容語の認識精度が改善された例を図 5 に示す．機能語の誤認識は改善されていないが，内容語の誤認識が改善されているため検索精度の改善を確認することができた．

以上の結果から，音声入力型情報検索システムに対するベイズリスク最小化音声認識の適用の有効性を示すことができた．

表 3　実験結果 2

	従来法	MBR 音声認識
認識精度(%)	68.7	68.9
検索精度(%)	72.0	72.9

入力発話　　：　理工学部 の 学費
従来法による認識　：　理工学部 の 学生
MBR による認識　：　理工学部 の 学費

入力発話　　：水曜の休講を教えて
従来法による認識　：水曜の休耕を　せ て
MBR による認識　：水曜の休講を　せ　て

図 5　提案手法により改善された例

6. おわりに

本研究では音声入力型大学情報検索システムに単語誤り率を最小化するベイズリスク最小化音声認識を導入した．評価実験を行なった結果，従来の尤度最大化音声認識では認識精度 68.7%，検索精度は 72.0%，ベイズリスク最小化音声認識では認識精度 68.9%，検索精度 72.9%という結果が得られ，音声認識精度ならびに検索精度を改善することが

Vol.2011-NL-201 No.5
Vol.2011-SLP-86 No.5
2011/5/16

でき提案手法の有効性を示すことができた．

今後は SVD 等による次元圧縮を行った検索の実現，より内容語を多く含んだ検索クエリでの実験を行う予定である．さらに，検索に重要な単語の認識精度の改善のため，重みつき単語誤り率の最小化音声認識の導入を行う[9].

謝 辞

本研究は科研費基盤研究(B)(21300066)の助成を受けたものである

参 考 文 献

[1] S.Furui,K.Maekawa, and H.Isahara, ”Toward the realization of spontaneous speech recognition – introduction of a Japanese priority program and preliminary results ” , Proc. ICSLP, Vol.3, pp 518-521, 2000.

[2] S.Furui, ”Recent advances in spontaneous speech recognition and understanding”, Proc. ISCA IEEE Workshop on Spontaneous Speech Processing and Recognition, pp. 1-6, 2003.

[3] 西崎博光，中川聖一, ”未知語を考慮したニュース音声記事の検索”，信学技報, NLC2001-77,SP2001-112, 2001.

[4] 秋葉友良，横田悠右，"認識候補から正解テキストへの翻訳に基づく講演音声ドキュメントのアドホック検索"，情報処理学会論文誌，Vol.50, No.2， pp.514-523 , 2009.

[5] 伊藤慶明，岩田耕平，石亀昌明，田中和世，李 時旭，"語彙制限のない音声文書検索における複数サブワードの統合―検索語彙に依存した検索性能推定指標の導入", 情報処理学会論文誌， Vol.50, No.2, pp.524-533, 2009.

[6] 宇野有，伊藤仁，伊藤彰則，牧野正三，”音声ドキュメント検索のための WWW を用いたインデクス改善”，第 4 回音声ドキュメント処理ワークショップ講演論文集, SDPWS2010-09.

[7] 西崎博光，”音声文書を対象とした音声入力型情報検索システムに関する研究”, 豊橋技術科学大学大学院博士論文，2003.

[8] 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，2010.

[9] 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”，信学論，Vol.J91-D No.5， pp1314-1324, 2008.

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

[11] A.Stolcke, Y.Konig, and M.Weintraub, “Explicit word error minimization in N-best list rescoring,” Proc.EUROSPEECH, pp.163-165, 1997.

[12] L.Mangu,E.Brill, and A.Stolcke, “Finding consensus in speech recognition: Word error minimization and other applicatiobns of confusion networks,” Comput.Speech Lang., vol.14, pp.373-400, 2000.

[13] G.Salton,M.McGill,”Introduction to Modern Information Retrieval”, McGraw-Hill, 1983.

[14] 松本裕治，北内啓，山下達雄，平野善隆，松田寛，高岡一馬，浅原 正幸，” 日本語形態素解析システム『茶筌』 version 2.2.1 使用説明書”, 2000.