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

2007－ＨＩ－122 (10)
2007－SLP－ 65 (10)
2007／2／10

文書検索型音声対話システムにおける ベイズリスクに基づく対話制御の最適化

翠 輝久　河原 達也

京都大学 情報学研究科 知能情報学専攻
〒606-8501 京都市左京区吉田本町
e-mail: misu@ar.media.kyoto-u.ac.jp

あらまし　自然言語テキストで記述された文書を検索・提示する音声対話システムにおける効率的な対話制御手法を提案する．このようなシステムでは，音声認識結果のN-best候補やコンテキスト情報を適切に利用することで，音声認識誤りや発話中の省略表現に対処できる可能性がある．また応答方法に関しても，確認の生成条件や情報の提示方法において，いくつかの選択肢が考えられる．本研究では，これらの生成可能な応答候補の中から最適なものを選択する過程を，情報提示の報酬と情報提示に直接関係しないターン数に基づくペナルティを用いて定義されるベイズリスクを最小化する枠組みとして定式化を行う．観光情報の検索・提示を行うシステム「京都版ダイアログナビ」において評価を行い，従来の単純な対話制御法と比較して，応答成功率と回答提示までに必要なターン数において大きな改善が得られた．

Bayes Risk-based Optimization of Dialog Management for Document Retrieval System with Speech Interface

Teruhisa Misu　Tatsuya Kawahara

School of Informatics, Kyoto University,
Kyoto 606-8501, Japan
e-mail: misu@ar.media.kyoto-u.ac.jp

Abstract　We propose an efficient dialogue management for an information navigation system based on a document knowledge base. It is expected that incorporation of appropriate N-best candidates of ASR and contextual information will improve the system performance. The system also has several choices in generating responses or confirmations. In this paper, this selection is optimized as minimization of Bayes risk based on reward for correct information presentation and penalty for redundant turns. We have evaluated this strategy with our spoken dialogue system "Speech Dialogue Navigator", which also has question-answering capability. Effectiveness of the proposed framework was confirmed in the success rate of retrieval and the average number of turns for information access.

1 はじめに

音声対話システムの研究対象は，関係データベースからWebのテキストや新聞記事などの一般的な文書へと広がりつつある [1, 2]．これらのシステムでは，ユーザ発話の音声認識結果と文書のマッチングが行われ，そのゆう度が高いものが提示される．このようなシステムは，博物館等の音声ガイドや車中での情報案内のような，音声が主要なモダリティである環境において特に有用である．我々はこのような文書検索タスクにおいて，検索要求のみではなく，特定の情報・事実を求める質問応答機能を利用することで，インタラクティブな情報提示を行う対話システム「京都版ダイアログナビ」を構築してきた [3, 4]．

円滑な音声対話を実現するためには，音声認識誤りに対する確認が不可欠である．しかし，不必要な確認はユーザに対してわずらわしいものであり，必要最小限の回数にとどめることが望ましいことから，従来より確認の効率化を行う対話制御に関する研究が行われてきた [5, 6]．しかし，これらは基本的に，SQLコマンドを用いたデータベース検索タスクを対象として定式化が行われている．そのため，「京都版ダイアログナビ」のような文書検索タスクに対して，これらの手法をそのまま適用することは難しい．

また本システムでは，ユーザの発話に対して検索要求・質問の判定を行い，それぞれに対応した応答を生成する．NTCIRなどの質問応答タスクとは異なり，実際のユーザの発話には検索要求か質問のいずれであるか判断が困難な場合も多く，また入力される質問の全てに対して，必ず回答が存在するとも限らない．そのため，ユーザの発話が質問である可能性が高い場合でも，質問の回答のみを提示するのが最適な応答であるとは限らない．

そこで本研究では，従来行われてきた対話制御方法を，質問応答を含む文書検索タスクに適用可能なものに拡張することを目指す．具体的には，ユーザが所望する情報を提示することによる報酬と，情報提示に直接関係のないターン数に基づいてペナルティを定義して，文書検索のゆう度や質問の回答生成の確信度も考慮し，ベイズリスクを最小化する枠組みを提案する．

U1：銀閣寺について教えてください．
S1：銀閣寺は京都市左京区にある東山文化を代表する臨済宗相国寺派の寺院です．銀閣寺の歴史やみどころについて何か知りたいことはありますか？
U2：いつ建てられましたか？
S2：銀閣寺の建立は1482年です．
U2：じゃあ，庭園を説明して．
……

図 1:「京都版ダイアログナビ」の対話例

2 情報検索システムにおける対話制御・応答生成

京都版ダイアログナビは，音声入力による文書検索システムであり，質問応答機能を利用することで，単純な情報提示に加えて，ユーザの聞き逃しや関連情報に対する要求などに対処することができる．このシステムにおける対話例を図1に示す．

2.1 解釈・応答生成において考慮する項目

我々は，このシステムを京都大学博物館の特別展示において，約3ヶ月間の運用を行った [3, 4]．その結果，応答生成に際に以下の点を改良することにより，システムの性能の改善を得られる可能性があることがわかった．

1. 音声認識のN-best候補の利用
音声認識誤りに対して頑健な対話を行うために，認識結果のN-best候補を利用することが有効である [7]．京都版ダイアログナビにおいても，音声認識のN-best候補中の全ての単語を用いて検索クエリを作成することにより，システム性能の改善が得られた．しかし逆に，ユーザが発話していない単語がクエリに含まれる可能性があり，その場合に検索性能の低下の原因となる．そのため，ユーザ発話の音声認識のN-best候補から最適な候補を選択して検索を行うことが，より望ましい [4]．また，これに関する先行研究として，秋葉ら [8] は，質問応答タスクにおける回答抽出スコアを考慮した選択（リスコアリング）手法を提案している．

2. コンテキスト情報の補完
一連の対話のコンテキストを考慮したマッチングを行うために，検索クエリに発話履歴中に含まれる単語を補完・追加している．しかし，この方法

は，ユーザが話題を変えた場合に，直前の話題のコンテキストを引きずってしまい，誤った検索を行う可能性がある．そのため，発話内容に応じてコンテキストを利用するか判断することが望ましい．

3. 確認・応答方法の決定

誤った内容の情報提示を避けるために音声認識やマッチングのゆう度が低い場合には，確認を行うことが望ましい．また，質問に対する回答抽出の確信度が低い場合には，質問の回答のみを提示するのではなく，文書全体を提示することが有効な場合もありうる．

本研究では，ベイズリスク最適化の枠組みの下でこれらの問題を統合的に扱い，最適な候補を選択・解釈し，応答生成を行う方法を提案する．

2.2 提案手法の概要

提案する対話制御は，検索クエリの生成方法や応答方法を変えることにより複数の応答候補を生成し，それらを比較・選択することにより実現される．

ここで，検索結果の文書 D を用いて生成する応答 $Act(D)$ は，以下の3つからなるものとする．一つ目は，文書の（確認なしでの）提示 $Pres(D)$ であり，文書 D を要約して応答を作成する．二つ目は文書 D を提示することに対する確認 $Conf(D)$ であり，文書のタイトルを基に「金閣寺でよろしいでしょうか？」といった確認を生成する．三つ目は，質問に対する回答の提示 $Ans(D)$ であり，文書 D を基に，ユーザの質問に対する回答を含む一文を提示する．

可能な応答候補に対して，それぞれ成功した場合の報酬と失敗した場合のペナルティ，及び成功する確率（＝信頼度により近似）に基づいて，ベイズリスクを定義し，これが最小になるものを選択する．

以上の手順の概要を以下にまとめる．また，処理の流れを図2に示す．

1. ユーザの発話の音声認識結果の第1・第2・第3候補，第1〜3候補をマージしたもの，それぞれにコンテキスト情報を加えたものの合計8個の検索クエリ $q_i(i=1,\ldots,8)$ を作成する．
2. 各クエリを用いて知識ベースの検索（マッチング）を行い，提示文書候補 D_i とその確信度 $P(D_i)$ を計算する．
3. 検索された文書 D_i から，文書の提示 $Pres(D_i)$，確認 $Conf(D_i)$，質問の回答として提示 $Ans(D_i)$ の3種類の応答候補を作成する．
4. 4(音声認識N-best候補)×2(コンテキスト情報の有無)× 3(応答方法)+1(リジェクト)=25の候補に対して，それぞれベイズリスクを計算し，これが最小となるものを選択してユーザに提示する．

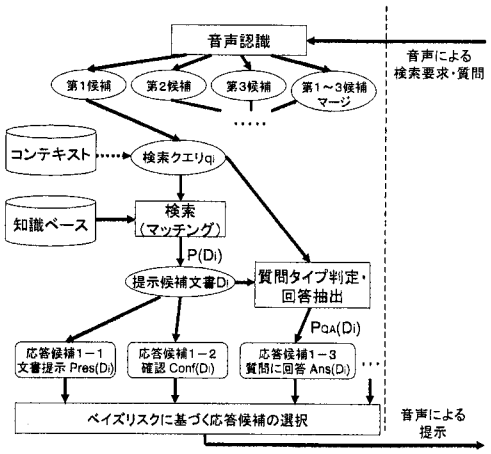

図 2: 提案手法の処理の概要

表 1: システムが利用する知識ベース

知識ベースの種類	件数	見出し（節）数	単語数
Wikipediaの京都に関する文書	269	678	約15万
京都市観光局・京都情報データベース	541	541	約7万
合計	810	1219	約22万

2.3 システムが利用する知識ベース

本研究では，システムが検索・提示する文書として，Wikipedia[1] の京都に関する文書と京都市産業観光局が提供する京都情報データベース[2] を使用する．それぞれの文書の概要を表1に，Wikipediaの文書の例を図3に示す．

これらの文書は，Webブラウザによりテキストの形で閲覧することを前提に作成されており，文体は書き言葉調であるため，音声合成によりそのまま読み上げるのは不自然になる．そのため，文末の助詞の変更，書き言葉特有の語彙の平易な表現への言い換えを行うことで話し言葉調に変換した．

[1] http://ja.wikipedia.org/
[2] http://raku.city.kyoto.jp/sight.phtml

慈照寺

(概要)
慈照寺は，京都府京都市左京区にあり，東山文化を代表する臨済宗相国寺派の寺院．通称銀閣寺，山号は東山．開基は，室町幕府 8 代将軍の足利義政，・・・

沿革

室町幕府 8 代将軍足利義政は，1473 年，嗣子足利義尚に将軍職を譲り，1482 年から，東山の月待山麓に東山山荘の造営を始めた．この地は，・・・

境内

錦鏡池を中心とする池泉回遊式庭園.「苔寺」の通称で知られる西芳寺庭園を模して造られたとされるが，江戸時代に改修されており，・・・

図 3: Wikipedia の文書の例

3 応答候補の生成方法

3.1 ベクトル空間モデルによる知識ベースの検索

ユーザの発話と知識ベース中の文書との類似度を計算するために，単語ベースのベクトル空間モデルを採用する．

ユーザ発話から作成する検索クエリベクトル q は，音声認識結果中の名詞に対して音声認識の信頼度で重み付けして作成する．コンテキスト情報を含むクエリは，現在のトピックに関する履歴中のユーザの発話中に含まれる単語も使用する．

知識ベースの文書も同様に，節を単位として，含まれる名詞に対して，タイトルに重み付けをした出現回数に基づく文書ベクトルを作成する．たとえば，図 3 の例においては，概要，沿革，境内の説明に対して，それぞれベクトルが作られる．

以上の手順で作成したクエリベクトル q と，文書ベクトル D を用いて内積類似度 $Product(q, D)$ を計算する．この類似度が最大の文書 D_i をベクトル q による検索結果とする．類似度を（0〜1 の値である）の確信度 $P(D)$ に変換するには以下のシグモイド関数を用いる．

$$P(D) = \frac{1}{1 + \alpha * \exp\{-1 * (Product(q, D) - \beta)\}}$$

ここで，α，β は定数である．

3.2 検索要求に対する応答の生成

検索された文書の提示応答 $Pres(D)$ は，文書 D の概要の提示を行うものである．ユーザの理解のしやすさを考慮して，文書中での文の出現位置と文間のつながりを手がかりに，重要文抽出による要約を行い，その結果を読み上げる．

また，確認 $Conf(D)$ は，文書のタイトルや見出しを利用して,「金閣寺でよろしいでしょうか？」のような応答文を生成する．

3.3 質問に対する回答の生成

質問に対する回答 $Ans(D)$ の生成は，質問タイプの判定と回答抽出の 2 つのプロセスから成る．

ユーザ発話の質問タイプの判定には，人手によるヒューリスティックなルールを用いる．たとえば，認識結果中に「誰ですか」という表現が含まれる場合には，人名をたずねる質問であると判定し,「いくらですか」という表現が含まれる場合には，金額をたずねる質問であると判定する．このように用意したルールにより，6 種類の質問タイプに対応する．

回答抽出には，テキストベースの質問応答システムで利用される一般的な手法を実装した．具体的には，文書 D の中に含まれる質問タイプに対応する固有表現 (NE) ごとに，以下の特徴量を用いて回答抽出スコアを計算して，それが最大となる NE を含む文を回答とする．

- NE を含む文にクエリ中の名詞が含まれる個数
- NE を含む文節に係る文節，NE を含む文節が係る文節にクエリ中の名詞が含まれる個数

なお，質問に対する回答の確信度 $P_{QA}(D)$ は，発話タイプ判定の根拠となる語句の音声認識の信頼度と，回答抽出のスコアを用いて計算する．

4 ベイズリスクに基づく応答候補の選択

次に，システムの応答を決定するための基準としてベイズリスクを定義する．これは，ユーザに所望の情報を提示したときにシステムが得る報酬と，確認や誤った情報を提示することによるペナルティに基づいて定義する．すなわち，ユーザが要求している情報を正しく提示した場合には，応答内容に応じた報酬を与える．逆に，誤った内容を提示した場合

や，候補のリジェクトを行った場合には，システムがその応答を行ったことにより，余計に費やすターン数（＝ユーザが次のクエリを発話するまでに必要な合計ターン数）に応じたペナルティを与える．ペナルティは正しい候補を提示した場合には0であるが，その他の場合には応答内容に応じた正の値をとる．たとえば，確認を行う場合には，[システムの確認＋ユーザの回答] の2ターン分，誤った情報を提示した場合には，[情報提示＋ユーザの訂正＋システムの謝罪＋再発話要求（または確認）＋ユーザによる再発話（または確認への回答）] の5ターン分のペナルティが与えられる．

システムが生成する各応答候補に対するベイズリスクは，文書検索の確信度 $P(D)$，ユーザの質問に対する回答の確信度 $P_{QA}(D)$，ターン毎のペナルティ $Penalty$，検索・質問応答成功時の報酬 Rwd_{Ret}，Rwd_{QA} ($Rwd_{Ret} < Rwd_{QA}$) を用いて以下のように記述できる．

- 文書 D を（確認なしで）提示

$$Risk(Pres(D)) = -Rwd_{Ret} * P(D) + (5 * Penalty) * (1 - P(D))$$

- 文書 D を提示することに対する確認

$$Risk(Conf(D)) = (-Rwd_{Ret} + 2 * Penalty) * P(D) + (3 * Penalty) * (1 - P(D))$$

- 文書 D を用いてユーザの質問に回答

$$Risk(Ans(D)) = -Rwd_{QA} * P_{QA}(D) * P(D) - \frac{1}{2} Rwd_{Ret} * (1 - P_{QA}(D)) * P(D) + 5 * Penalty * (1 - P(D))$$

- リジェクト

$$Risk(Reject) = 2 * Penalty$$

なお，質問に対する回答を提示すべき場合に，検索結果を提示した場合には，通常の半分の報酬を与える．これは，質問の回答を直接提示することに失敗した場合であっても，文書の要約を提示することにより，質問の回答が含まれる場合があるためである．

$Penalty = 1$，$Rwd_{Ret} = 4$，$Rwd_{QA} = 8$ とした場合のベイズリスクの計算例を図4に示す．

・音声認識結果の第1候補:「銀閣寺が知りたい」
→ 銀閣寺 P(銀閣寺) = 0.4
- $Risk(Pres$(銀閣寺)$) = -1.6 + 3.0 = 1.4$
- $Risk(Conf$(銀閣寺)$) = -0.8 + 1.8 = 1.0$

・音声認識結果の第2候補:「金閣寺が知りたい」
→ 金閣寺 P(金閣寺) = 0.2
- $Risk(Pres$(金閣寺)$) = -0.8 + 4.0 = 3.2$
- $Risk(Conf$(金閣寺)$) = -0.4 + 2.4 = 2.0$

...

・リジェクト
- $Risk(Reject) = 2.0$

↓

応答内容：銀閣寺の文書を提示することの確認:「銀閣寺でよろしいでしょうか？」

図 4: ベイズリスク計算の例

5 提案手法の評価

提案する対話戦略を評価するために，京都大学総合博物館企画展「コンピュータに感覚を」（2006年6月～8月）の展示システム「京都版ダイアログナビ」において収集されたユーザの発話データを用いる．企画展前半の30日間で収集された，ドメイン内の検索要求・質問1416発話 (検索要求1084発話，質問332発話) を人手により書き起こし，回答となる文書・NEを付与した．

これらの発話に対するシステムの応答成功率と，ユーザが所望の情報を得るために必要なターン数により評価を行った．なお，ユーザ発話が検索要求である場合には，回答となる文書を提示した（確認した）場合に応答成功，質問に対しては，回答となる語句を含む応答を行った場合に応答成功とした．また，回答提示までに必要なターン数はユーザが再発話を行った際にシステムが正しい回答を提示できる確率を60%[3]とし，ユーザの所望の情報を提示するまでに必要なターン数の期待値を求めることにより計算した．また，$Penalty$・Rwd の値は，クロスバリデーションにより決定した．この結果を表2に示す．

比較対象として，以下の従来法1，従来法2により応答を生成した結果との比較を行った．なお，従来法1がベースラインシステム，従来法2が京都大学博物館で運用を行ったシステムと同等のものである．

[3] ベースラインの平均応答成功率を基に設定

表 2: 提案手法による応答生成結果

	応答成功率	回答提示までに必要なターン数
検索	67.4%	4.35
質問応答	57.8%	4.95
合計	65.2%	4.49

表 3: 従来手法との比較

	応答成功率	回答提示までに必要なターン数
従来法 1 (baseline)	59.2%	5.44
従来法 2	63.4%	5.07
提案手法	65.2%	4.49

従来法 1（ベースラインシステム）

- 音声認識結果の第 1 候補のみを用いて検索クエリを作成
- 現在のトピックに関するコンテキスト情報を補完
- 認識結果中の名詞の音声認識信頼度が低い場合に確認を生成
- 質問タイプ判定器により，ユーザ発話が質問であると判定された場合には，質問の回答を提示

従来法 2（博物館システム）

- 音声認識結果の第 1〜3 候補中の全ての単語を用いて検索クエリを作成
- 他の条件は従来法 1 と同様

比較結果を表 3 に示す．提案手法による性能の改善は，ベースライン手法と比較して応答成功率で 6%，回答提示までに必要なターン数で約 1 ターンである．また，提案手法によって選択された応答候補を表 4 にまとめる．多くの応答が音声認識結果の第 1〜3 候補を単独で用いる検索クエリから生成されていることがわかる．また，質問への回答はコンテキストを含むクエリから生成されることが多く，主語等の省略・照応が頻繁に行われがちな質問に対して，コンテキストの適切な補完が行われていることがわかる．

6 おわりに

音声入力により情報検索・質問応答を行う対話システムにおいて，ベイズリスクに基づいて最適な応答を生成する枠組みを提案した．1416 発話による評価の結果，従来手法と比較して，少ないターン数で高い応答成功率を得られることを確認した．

本研究では，ベイズリスクの計算にターン数を基にしたペナルティを用いているが，ターンの質（内容）について考慮をしていない．しかし実際の対話では，同じ 1 ターンであっても，ターンの質が同じわけではない．例えば，確認と文書の提示では，応答にかかる時間もユーザに伝わる情報量も大きく異なる．そのため，今後このようなターンの質を報酬やペナルティに反映させることの検討を行っていく予定である．

表 4: 提案手法により選択された応答候補

	コンテキストなし			コンテキスト使用		
	検索	確認	QA	検索	確認	QA
第 1 候補	260	108	62	0	129	2
第 2 候補	146	38	29	0	1	6
第 3 候補	222	38	44	2	2	5
merge	78	8	3	41	0	90
reject	50					

参考文献

[1] A. Fujii and K. Itou. Building a test collection for speech-driven Web retrieval. In *Proc. Eurospeech*, 2003.

[2] 西崎博光, 中川聖一. 音声認識誤りと未知語に頑健な音声文書検索手法. 電子情報通信学会論文誌, Vol. J86-DII, No. 10, pp. 1369–1381, 2003.

[3] 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-62-13, 2006.

[4] 翠輝久, 河原達也. 質問応答技術を利用したインタラクティブな音声対話システム. 人工知能学会研究会資料, SIG-SLUD-A602-6, 2006.

[5] 新美康永, 小林豊. 音声認識の誤りを考慮した対話制御方式のモデル化. 情報処理学会研究報告, 95-SLP-5-7, 1995.

[6] 堂坂浩二, 安田宜仁, 相川清明. システム知識制限下での効率的対話制御法. 自然言語処理, Vol. 9, No. 1, pp. 43–63, 2002.

[7] 翠輝久, 駒谷和範, 清田陽司, 河原達也. 音声対話によるソフトウェアサポートタスクのための効率的な確認戦略. 電子情報通信学会論文誌, Vol. J88-DII, No. 3, pp. 499–508, 2005.

[8] T. Akiba and H. Abe. Exploiting Passage Retrieval for N-Best Rescoring of Spoken Questions. In *Proc. Interspeech*, 2005.