2003－SLP－47　(11)
2003／7／19

音声対話によるソフトウェアサポートタスクのための確認戦略

翠 輝久　駒谷 和範　河原 達也　奥乃 博　木戸 冬子†

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
† マイクロソフト株式会社
e-mail: misu@ar.media.kyoto-u.ac.jp

あらまし　汎用的な大規模知識ベースを自然言語音声で検索するタスクにおいては，話し言葉特有の冗長性や音声認識誤りに対処する必要がある．本研究では，ユーザ発話の認識結果から検索に有用な部分を同定し，確認を行う手法を提案する．まず、検索に決定的な影響を与える箇所について，認識誤りやドメイン外である可能性が高い場合は検索前に確認を行う．この際には，検索に用いる知識ベースのみから作成した言語モデルで計算した検索整合度を利用する．次に、結果として検索に影響を与える箇所について，音声認識結果の N-best 候補を実際に検索した結果の違いに基づいて検索後に確認を行う．以上の対話戦略をソフトウェアサポートを行うダイアログナビのフロントエンドとして実装した．評価の結果，単純に音声認識結果を用いる場合より検索成功率が向上し，また認識の信頼度を用いる確認戦略よりも効率的に確認が行えることを示す．

Confirmation Strategies for Software Support with Spoken Dialogue

Teruhisa Misu　Kazunori Komatani　Tatsuya Kawahara
Hiroshi G. Okuno　Fuyuko Kido†

School of Informatics, Kyoto University, Kyoto 606-8501, Japan
† Microsoft Co., Ltd.

Abstract　In most of existing spoken dialogue systems, a set of keywords to achieve the task achievement are defined beforehand. Thus, interpretation of utterances and confirmation strategy can be realized by focusing on such keywords. When the task is information retrieval from large-scale text knowledge base, however, it is impossible neither to define such keywords explicitly, nor to make confirmation by focusing on such keywords. We propose efficient confirmation strategies: confirming portions that critically affect the retrieval beforehand, and confirming portions that have some effect on the actual retrieved results afterwards. The method are evaluated by the retrieval success rate using 235 queries uttered by 10 users. The experimental result shows that the proposed confirmation strategy achieves a better success rate with less numbers of confirmation.

1 はじめに

自然な音声で情報検索を行える音声対話システムが実用化されつつある [1][2]．音声対話システムにおいて，発話からユーザの意図を解釈する手続きは不可欠である．バスの運行情報案内タスク [3] などスロットフィリング型のタスクでは，発話の中から検索に必要なキーワードを抽出することでユーザの意図を解釈し，それが同定できなければ確認するといった方法論を用いることができた．しかし，マニュアル [4] や Web ページなどのテキストで記述された大規模知識ベースを検索する際には，キーワードの集合を明確に定義することが不可能なため，発話を自然言語として解釈する必要がある [5]．

しかし，音声で自然言語を入力する場合に，単純に音声認識結果をそのまま用いて検索をすればよいわけではない．この原因の一つに，まず音声認識誤りがある．キーワードが明確に定義されている場合は，それらを 1 つずつ確認することができるが，一般的な自然言語入力の場合は難しい．もう一つの原因として，正確な書き起こしが得られた場合でも，音声にはフィラーや多様な文末表現など，冗長性が多いため，発話のすべてが検索に必要なものとは限らないという問題がある．したがって，検索に必要でない部分に対してはその部分を取り除いて検索したり，逐一確認を行わないようにすることが望ましい．このため，音声認識結果から検索に有用な部分を自動的に判別する枠組が必要になる．

本研究では，検索に用いる知識ベースのみから学習した統計的言語モデルによる尤度と，音声認識の N-best 候補に対する検索結果の両方を用いて，音声認識結果の各文節が検索に必要かどうかを判定する．音声認識の言語モデルと検索文書の言語モデルを使い分けることにより，認識の頑健性を向上させながら検索に必要でない部分を検出する．さらに複数の候補を求めて検索結果を得ることで，検索結果に違いを与える部分を同定する．これらの情報を用いて検索において重要な部分に対して効率的に確認を行う手法を提案する．

2 音声言語による大規模知識ベースの検索システム

本研究では，対象タスクドメインとしてマイクロソフト社のソフトウェアサポート用知識ベースを用いる．この知識ベースのテキスト集合は以下の 3 つから構成されている．

- 用語集
- ヘルプ集
- サポート技術情報

これらのテキスト集合のうち，用語集，ヘルプ集は見出し語とその説明からなり，サポート技術情報は見出し，概要（現象），詳細情報から構成される．サポート技術情報の例を図 1 に示す．これらすべてが例のように自然言語によって記述されている．さらに，表 1 のように知識ベースが大規模であることも特徴の一つである．

[HOWTO] Windows XP で音声認識を使用する方法

この資料は以下の製品について記述したものです．

- Microsoft Windows XP Professional
- Microsoft Windows XP Home Edition

概要 この資料では，Windows XP で音声認識を使用する方法について説明しています．Microsoft Office XP の音声認識をインストールしているか，または，Office XP がインストールされたコンピュータを新たに購入した場合は，すべての Office アプリケーションや，音声認識が利用可能なその他のアプリケーションで音声認識を使用できます．

詳細 音声認識は，音声をテキストに変換するオペレーティング システムの機能です．音声認識エンジンと呼ばれる内部ドライバによって，単語が認識され，テキストに変換されます．音声認識エンジンは，‥

図 1: マイクロソフト社ソフトウェアサポート用知識ベースの例

表 1: マイクロソフトソフトウェアサポート用知識ベース

知識ベースの種類	件数	文字数
用語集	4707	約 70 万
ヘルプ集	11306	約 600 万
サポート技術情報	23323	約 2200 万

これまで東京大学で，ユーザのテキスト入力文に対して知識ベースを検索する質問応答システムとしてダイアログナビ [6] が開発されている．ダイアログナビの特徴として，自然言語入力文と知識ベースを柔軟にマッチングするために，係り受け関係や同義表現を考慮して解釈していることが挙げられる．すなわち，自立語の一致だけではなく，木構造の文節の深さの一致や，係りタイプ等も評価し，それを正規化したスコアをマッチングの尺度として用いている [7]．この際，用語集の見出し，ヘルプ集のタイトル，サポート技術情報の文章全体を，入力文とのマッチングの対象としている．

本研究では，バックエンドとしてダイアログナビを使用して，音声入力により検索するシステムを作

成する．この際に問題となる認識誤り，ドメイン外発話といった音声言語の問題に対して頑健にユーザ発話の理解を行うための確認戦略を提案する．

3 検索整合度と検索重要度を用いた確認戦略

認識誤りの可能性が高い部分全てを一つずつ確認するのは非効率的である．また，認識誤り箇所が常に検索に悪影響を与えるとは限らない．そこで，本研究では単語ごとの認識誤りによる損失を考慮して，確認の方法を切り替える．検索に与える影響が常に大きい語句は，検索を実行する前にユーザに確認する．次に，常に決定的な影響を与えるわけではないが，結果として検索結果にある程度の影響を与える箇所を同定し，検索後に確認する．この検索に与える影響の度合の指標として検索重要度を定義する．

またこれとは別に，入力において認識誤りやドメイン外発話により，検索文として適当でない場合があるので，この指標として検索整合度を導入する．これは，事前確認を行うかの判断やマッチングの際の重みとして用いる．

以上の確認を組み込んだシステムの処理の流れは以下の通りである．

1. ユーザの発話を音声認識する．
2. 認識結果に対して検索整合度を計算する．
3. 検索整合度が低い重要語句を確認する．
4. 認識結果の複数候補を用いてダイアログナビで検索する．
5. 検索重要度を計算し，それが高い場合には確認対話を生成する．
6. 検索結果をユーザに提示する．

全体の処理の流れを図 2 に示す．

3.1 検索整合度の計算

検索整合度の計算には，検索対象である知識ベースのみから学習した言語モデルによる単語パープレキシティを使用する．これは，検索対象との整合性を示す尺度である．音声認識結果中の認識誤りである箇所は文脈的に不自然である場合が多く，また検索に直接関係がない語句は知識ベース内での出現確率が低いためパープレキシティは高くなる．このように，音声認識時と異なる言語モデルによりパープレキシティを計算することで，認識誤り箇所や，認識したが検索には重要でない箇所を同時に検出できる．パープレキシティを検索整合度 (relevance score) に変換するには以下の関数を使用する．

$$RS = \frac{1}{1 + \exp(\alpha * (\log PP - \beta))}$$

本研究では，部分的な認識誤りを棄却するために

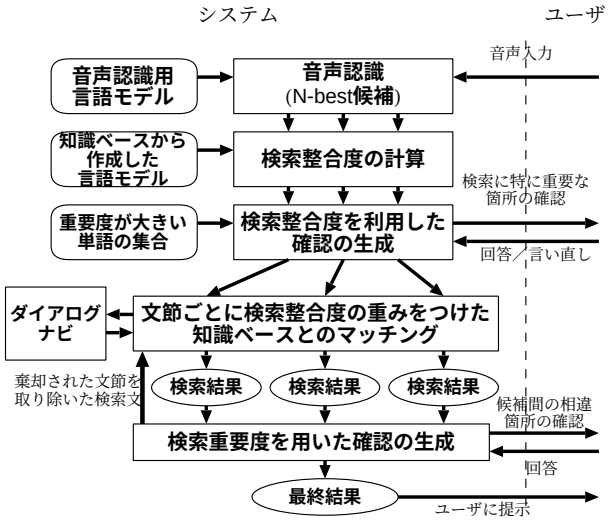

図 2: 確認対話戦略を導入した検索システムの概要

文節単位で検索整合度を計算する．その手順を以下に示す．

1. 認識結果を構文解析ツール KNP[8] を用いて文節単位に区切る．
2. 区切られた文節にそのコンテクストとして前後 1 単語を付け加える．
3. 知識ベースのみから作成した言語モデルでパープレキシティを計算する．
4. パープレキシティを検索整合度に変換する．

この手順を，実際の認識結果に適用した例を図 3 に示す．この例では，文頭の「新しく買ってきた」という検索に直接関係がない文節と，文末を誤認識した「以降」の文節のパープレキシティが高くなっている．

3.2 検索整合度を用いた重要語句の確認

検索に決定的な影響を与える (=検索重要度がきわめて大きい) と予測される語句が誤認識された場合，検索が失敗する可能性が高い．そのため，こうした語句は検索を実行する前にユーザに確認する．ソフトウェアサポートタスクでは，プロダクト名がこれにあたる．現時点では，プロダクト名のリストは人手により事前に用意した．

検索適合度が閾値 θ 以下である文節にプロダクト名が含まれる場合には，ユーザに認識結果を提示し，確認する．ユーザは提示された文節が認識誤りであると判断した場合には，その文節を認識結果から取り除くか，その文節のみを言い直すかを選択できる．こうして得られた結果を次のマッチングモジュールに渡す．

ユーザ発話：「新しく買ったＸＰのパソコンでＦＡＸ機能を使うにはどうしたらいいですか？」

音声認識結果「新しく買ったＸＰのパソコンでＦＡＸ機能を使うにそのA以降」

構文解析により文節単位に分割：「新しく／買った／ＸＰの／パソコンで／ＦＡＸ機能を／使うに／その／Ａ／以降」

前後 1 形態素を追加し，パープレキシティを計算：

<S>**新しく**買った	$PP =$	499.57
新しく**買った**ＸＰの	$PP =$	2079.83
買った**ＸＰの**パソコン	$PP =$	105.64
の**パソコンで**ＦＡＸ	$PP =$	185.92
で**ＦＡＸ機能を**使う	$PP =$	236.23
を**使うに**その	$PP =$	98.40
に**その**Ａ	$PP =$	1378.72
その**Ａ**以降	$PP =$	144.58
Ａ**以降**</S>	$PP =$	27150.00

<S>，</S>はそれぞれ始端記号，終端記号

図 3: 検索整合度 (パープレキシティ) の計算例

3.3 検索整合度を用いた重み付きマッチング

検索整合度が低い文節は，認識誤りであるか，検索に直接関係ない可能性が高い．そのため，知識ベースとマッチングを行う際に各文節の検索整合度 RS をその文節に対する重みとして用いる．これにより，認識誤りや発話中の冗長部分による検索への影響を軽減できる．

3.4 検索重要度を利用した確認戦略

ユーザの発話が短く，比較的明瞭に発声されている場合には，認識結果の第 1 候補が誤りであっても，N-best 候補の中に正しい認識結果が含まれる可能性がある．しかし，検索に影響が少ない付属語などの置換も多いため，これら全てを確認するのは非効率的である．

そこで，音声認識結果の N-best 候補それぞれに対する検索結果を用いて検索重要度を定義する．その上で，結果的に検索に違いを与えた部分に対して確認を行う．

まず，音声認識結果の N-best 候補間の相違箇所を同定する．次に，この N-best 候補それぞれについて実際に検索を行い，結果が異なる場合にその相違の大きさを検索重要度 (significance score) として定義する．ここでは検索重要度は，検索の結果得られるスコアの高い 5 候補中で，異なる候補の割合とする．

具体例を図 4 に示す．認識結果の第 1 候補と第 2 候補では「数式」と「数字」の部分が異なる．この時，各々の検索結果の上位 5 候補中，4 候補が異なるため,「数式」と「数字」の部分の検索重要度は 0.8 となる．

第１候補 ＷＯＲＤ２００２で**＜数字＞**を入力する方法を教えてください
1. Word で現在の日付と時刻を入力する
2. WORD:日本語と英数字の間のスペースを取る
3. WORD:入力した文字の書式を確認する
4. WORD:手書きのサインを入力する
5. WORD:文字の背景に透かし文字を入れる

第２候補 ＷＯＲＤ２００２で**＜数式＞**を入力する方法を教えてください
1. Word で数式を挿入する
2. Word で現在の日付と時刻を入力する
3. スプレッドシートで数式を入力する
4. PowerPoint で数式を挿入する
5. Excel で数式を入力する

$$SS = 4/5 = 0.8$$

図 4: 検索重要度 SS の計算例

検索重要度が閾値を越えている場合には，その相違部分をユーザに提示し確認する．なお，今回音声認識の際に出力する候補数 N は 3 とし，確認を行う閾値は 0.5 とした．ユーザが提示された候補の中から適切なものを選択すると，対応する検索結果が表示される．提示された候補が全て適当でなかった場合には，その文節を棄却して再検索した結果をユーザに提示することもできる．検索重要度が閾値以下の場合には確認を行わず，第 1 候補による検索結果をそのまま表示する．

4 実装と評価実験

提案手法の評価のために，マイクロソフト社の Web ブラウザ Internet Explorer 6.0 上で動作するシステムを実装した．音声認識は，クライアント PC の Julius for SAPI[1] [9] により行う．また，ユーザに対する確認は画面に出力し，ユーザは選択肢ボタンをクリックすることによりシステムからの質問に回答する．

4.1 ドメインに対応した言語モデルの作成

音声認識用の言語モデルの作成に用いる学習データとして以下のコーパスを用いる．表 2 にこれらの概要を示す．

- 「マイクロソフトサポート技術情報」の「タイトル・概要」「現象・症状」部分
- 「マイクロソフト話し言葉検索 [10]」に寄せられた質問集

[1] http://julius.sourceforge.jp/sapi/

- ダイアログナビに寄せられた質問集
- ソフトウェアサポートの疑似対話を書き起こしたもののユーザ発話部分

これらのコーパスから統計的言語モデルを作成する際，Word や Excel といったアプリケーション名を APPLICATION というクラスで扱う．同様に定義したクラス数は合計 10，クラス内のエントリ数の平均は約 15 である．

表 2 に示すように，想定されるユーザの発話にスタイルが最も近いソフトウェアサポートの疑似対話の書き起こしが，他のコーパスと比べて極端に少ない．したがって，前者に重みをつけて混合した [11] ものを．音声認識用の言語モデルとして用いる．

なお，検索整合度を計算するための言語モデルは，「マイクロソフトサポート技術情報」のコーパスのみを用いて作成した．

4.2 音声を用いた検索システムの評価

評価用データは本システムを利用したことのない 10 名 (男 7，女 3) の被験者により収集した．設定した想定場面に基づいて各 11 課題，これとは別に，自由に 3 課題について検索を行ってもらった．ただし，検索結果として，ふさわしい候補が提示されない場合には各課題につき 3 度まで言い直しを許した．また，ユーザにはシステムが話し言葉による発話も理解できることを伝えた．その結果，合計 139 課題，235 発話を得た．10 人の発話の音声認識率は平均で 65.6%である．

これらの収集した結果に対して，以下の 3 つの条件で検索実験を行った．

1. ユーザ発話の正確な書き起こし (人手で作成) を用いて検索した場合 [書き起こし入力]
2. 音声認識結果の第 1 候補を用いて検索した場合 [認識結果入力]
3. 検索整合度と検索重要度の両方を用いて検索を行い，生成する確認に対してユーザが適切に応答した場合 [提案手法]

これらの条件での検索の成功率を表 3 に示す．なお，ユーザに最終的に提示した候補の中に最初の質問の答えとなる候補が含まれていた場合に検索成功とした．全 235 発話の成功率では，提案手法により検索を行った場合は音声認識結果の第 1 候補をそのまま用いて検索を行った場合よりも大幅に検索の成功率が上昇している．また，各課題のユーザの最終発話のみを比較した場合には，提案手法での成功率は，発話の正確な書き起こしを利用して検索を行った場合に近づいている．ユーザ発話の例を図 5 に示す．

次に，システムが生成した確認に関して検証を行った．生成された確認の回数は 126 回である．これは，1 課題あたりおおよそ確認が 1 回行われていること

- 検索が成功したもの
  - OUTLOOK で画像を送るにはどうしたらいいですか？
  - どうやって DVD を見るんですか？
  - WINDOWS2000 で，えーっと，パソコンを起動して，ログインしようとした時に，え，パスワードを忘れてしまったんですが．
  - と，OUTLOOK2002 を利用していますが，メールがいっぱい溜ってきて，重要なメールデータも多いのでバックアップしたいと思っています．
  - デジカメで撮った写真を友だちにメールで送りたいんですが，えー，メールソフトには普段通り OUTLOOK を使いたいのですが，画像を送ったことがないので，どうしたらいいかわかりません．
- 検索が失敗したもの
  - ウィルスが怖いんですが，どうしたらいいんですか？
  - えっと，WINDOWS で壁紙を変更しても，その前の壁紙が起動時に一瞬だけ表示されるんですが．
  - と，インターネットをしていて，あるホームページを見ていたら，えーっと，表示されるときに JAVASCRIPT を ON にしてくださいという表示が出ました．

図 5: ユーザ発話の例

になる．このうち，検索整合度を用いた事前確認の回数は 51 回あり，検索重要度を用いた事後確認が 75 回であった．検索整合度を用いた確認により，確認を行わない場合と比べて 9 発話で検索が成功した．同様に，検索重要度を用いた確認により，6 発話で検索成功が増えた．

提案手法の確認回数を評価するために，音声認識結果の N-best 候補から計算される信頼度 [12] を用いた確認を行う場合との確認回数，検索成功率を比較した．確認を行うための信頼度の閾値θ_1として，0.4，0.6，0.8 の 3 通りを用いた．信頼度が閾値以下の自立語を確認するものとし，それが誤認識されたものであった場合には，その単語を含む文節を棄却して検索した．

この結果を表 4 にまとめる．提案手法は，音声認識の信頼度の閾値θ_1が 0.8 の場合に比べて確認回数を半分以下に抑えながら．高い検索成功率を得ている．以上より，提案手法の確認が効率的であることが確かめられた．

5 おわりに

本研究では，ソフトウェアサポートタスク用の知識ベースを対象とした検索タスクにおいてユーザの多様な音声発話に対して頑健に検索を行うための確認戦略を提案した．

音声認識結果に対して検索整合度と検索重要度の

表 2: 言語モデル作成のための学習データの内容

	入力	文体	フィラー	文の数	のべ単語数	語彙サイズ
サポート技術情報	テキスト	書き言葉	なし	22,133	682,937	11,628
話し言葉検索	テキスト	話し言葉	なし	137,703	1,744,582	19,586
ダイアログナビ	テキスト	話し言葉	なし	16,541	184,310	8,223
疑似対話の書き起こし	音声	話し言葉	あり	88	4,864	896

表 3: 検索成功率

評価対象	発話数	書き起こし入力	認識結果入力	提案手法
全発話	235	184(78.3%)	104(44.2%)	134(57.0%)
各課題の最終発話	139	125(89.9%)	93(66.9%)	118(85.0%)

表 4: 音声認識の信頼度を用いた確認戦略との確認回数，検索成功率の比較

	提案手法	信頼度 ($\theta_1 = 0.4$)	信頼度 ($\theta_1 = 0.6$)	信頼度 ($\theta_1 = 0.8$)
確認回数	126	76	173	284
検索成功率	134(57.0%)	113(48.1%)	125(53.2%)	129(54.9%)

2 つの尺度を導入し，検索に決定的な影響を与える箇所は検索を実行する前に確認し，結果として検索に影響を及ぼす箇所は検索結果の違いに基づき確認を行う戦略を考案した．

提案する対話戦略の有効性を評価するために，東京大学で開発されたダイアログナビをバックエンドにした，音声入力フロントエンドを実装した．評価実験の結果，確認回数を削減しながら，高い検索成功率を得られることが示された．

謝辞

本研究は，東京大学の黒橋禎夫助教授，清田陽司氏との共同研究である．両氏の貴重な貢献に感謝します．

参考文献

[1] 藤井敦. 音声による言語バリアフリーな他言語情報アクセス. 情報処理学会研究報告, SLP-44-33, 2002.

[2] S. Harabagiu, D. Moldovan, and J. Picone. Open-domain voice-activated question answering. pp. 502–508, 2002.

[3] 安達史博, 河原達也, 奥乃博, 岡本隆志, 中嶋宏. Voice XML の動的生成に基づく自然言語音声対話システム. 情報処理学会研究報告, 2002-SLP-40-23, 2002.

[4] 伊藤亮介, 駒谷和範, 河原達也. 機器操作マニュアルの知識と構造を利用した音声対話ヘルプシステム. 情報処理学会論文誌, 第 43 巻, pp. 2147–2154, 2002.

[5] 駒谷和範, 河原達也, 清田陽司, 黒橋禎夫, Pascale Fung. 柔軟な言語モデルとマッチングを用いた音声によるレストラン検索システム. 情報処理学会研究報告, 2001-SLP-39-30, 2001.

[6] ダイアログナビ. http://www.microsoft.com/japan/navigator/.

[7] 清田陽司, 黒橋禎夫, 木戸冬子. 大規模テキスト知識ベースに基づく自動質問応答–ダイア ログナビ–. 言語処理学会第 8 回年次大会発表論文集, pp. 271–274, 2002.

[8] 黒橋禎夫, 長尾真. 並列構造の検出に基づく長い日本語文の構文解析. 自然言語処理, Vol. 1, No. 1, pp. 35–57, 1994.

[9] 住吉貴志, 李晃伸, 河原達也. 音声認識エンジン Julius/Julian の API 実装. 情報処理学会研究報告, 2001-SLP-37-16, 2001.

[10] マイクロソフト話し言葉検索. http://www.microsoft.com/japan/enable/nlsearch/.

[11] 長友健太郎, 西村竜一, 小松久美子, 黒田由香, 李晃伸, 猿渡洋, 鹿野清宏. 相補的バックオフを用いた言語モデル融合ツールの構築. 情報処理学会研究報告, 2001-SLP-35-9, 2001.

[12] 駒谷和範, 河原達也. 音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理. 情報処理学会論文誌, 第 43 巻, pp. 3078–3086, 2002.