

情報検索タスクにおける音声対話システムの対話管理手法の考察

井ノ上直己[†] 谷戸文廣[†] 橋本和夫[†]

[†]KDD 研究所

〒 356 上福岡市大原 2-1-15

E-mail: {inoue, yato, kh}@lab.kdd.co.jp

本稿では、新聞記事検索のような情報検索タスクにおける音声対話システムの対話制御方式について検討を行う。対話の特徴を明らかにする目的で、情報検索システムを作成し、Wizard of OZ 方式で利用者とシステムとの対話の収集・分析を行った。その結果、検索結果が空の場合の処理が必要であること、利用者が必要な記事が複数ある場合の対話方法を工夫する必要があること、さらに利用者の2回目以降の対話における入力キーワード数が少ないため、もっと多くのキーワードを入力できるように対話を制御する必要があること、が示された。

An Investigation of Dialogue Management Method for a Spoken Dialogue System in Information Retrieval Task

Naomi INOUE[†] Fumihiro YATO[†] Kazuo HASHIMOTO[†]

[†]KDD Research and Development Laboratories

(2-1-15, Ohara, Kamifukuoka-shi, 356 JAPAN)

E-mail: {inoue, yato, kh}@lab.kdd.co.jp

In this paper, a spoken dialogue management method for information retrieval task is investigated. Dialogues to retrieve articles of newspapers are collected by using wizard of OZ method, and analyzed. As a result, the response from the system should be controlled to work well when both no article and many articles are retrieved. Moreover, the response should be controlled to get more information (keywords) from a user.

1 はじめに

インターネットが普及し種々の情報が容易に手に入る環境が整ってきた。それに伴い、利用者が知りたい内容を適切に検索できる情報検索システムの研究がさかんに行われるようになってきている。しかし、高齢

者や障害者はキーボード操作に不慣れであったり、障害のためにキーボード入力ができなかったりするため、情報検索システムを利用できない。そこで、音声による検索キーワード入力等が行えると、その利用の可能性は飛躍的に広がるものと考えられ、我々は音声対話

機能を有する情報検索システムの開発を目標として研究を進めている。

本稿では、新聞記事の検索を対象とした情報検索タスクへの音声対話方式を検討するため、まず、関連する分野の従来からの研究を概観し、次に、収集した利用者とシステムとの対話の分析結果について報告する。

2 関連分野の従来からの研究

人間とコンピュータとが対話を行うシステムに関する研究は、古くから行われており、さまざまな研究成果が報告されている。本章では、まずこれらの成果および課題について概観する。

2.1 質問応答システム

自然言語入力により利用者とシステムが対話を行いながら、データベース検索を行うことを目的とした対話システムは、古くから質問応答(question answering)システムとして、自然言語理解や自然言語生成の問題として研究されてきた。その成果として、プラン認識[1]、協調的な応答[2]、談話構造[3]などの重要な枠組みが提案されている。我が国でも、質問応答システムに関する対話処理技術の研究は種々の観点から行われており、例えば、質問応答における意図の把握や話題の管理手法の研究[4]、一般の会話が守らなければならない会話の原則として、Griceの会話の原則[5]があり、システムが生成する応答文にもこの原則を適用することでシステムが利用者と自然な会話を行おうとする研究[6]、データベース検索結果が空であった場合、単に検索に失敗した事実だけを応答するのではなく、代案を通知したり失敗した原因を通知するなどの利用者との対話を協調的に行うための研究[7][8]、などがある。これらの研究は、利用者とシステムとの対話を円滑に行うためには、システム内部における対話の管理方法だけでなく、

(1) 自然な対話のための応答表現の生成、

(2) 検索結果が空であった場合の利用者への質問内容の選定、

も重要な課題であることを示している。

2.2 音声対話システム

先に示したような研究成果にも関わらず、実用レベルに達している質問応答システムはほとんどないと思われる。従来システムでは、キーボード入力による対話が前提であり、キー入力数が多いなどの理由で会話が持つ即時性の特徴が十分いかせなかったことが原因の一つだと考えられる。

近年の音声認識技術の目覚ましい進歩に基づき、従来のキーボード入力に代わって、音声入力によりデータベース検索を行う対話システムの研究が活発になってきている[9][10]。音声入力による対話システムでは、キーボード入力にくらべて情報伝達速度は速いという長所はあるが、内容の確認のための対話が頻繁におこるようでは、やはり利用者には煩わしいものとなり、受け入れられない可能性がある[11]。そのため、キーボード入力から音声入力へ対話システムを発展させた場合、最も重要となる課題は、

(3) 認識誤りへの対処

であると言える。この問題に対処するため、認識誤りを含んでいることを前提とした対話処理手法の研究も行われている[12][13]。

2.3 情報検索システム

新聞記事のような情報検索システムでは、現状では入力キーワードとのexact matchが主流であるが、これでは十分な精度が得られないという指摘がある[14]。そのため、検索精度の向上のため、exact matchに対してpartial matchの手法の研究がさかんに行われている。partial matchの手法としては、VSM(Vector Space Model)[15]を用いた手法が代表的であり、VSMでは、検索要求(query)と文書(document)の両方を、語の重

みを要素としたベクトルとして表わし、このベクトル間の類似度を計算して類似度の高い文書を求める手法である。この手法では、文書を表現するベクトルの要素は互いに独立していることが仮定されているが、ベクトルの要素として語を選んだ場合、実際には要素間に何らかの関連性があるため、この仮定は成り立たない。そこで、ベクトルの要素として語を選ぶのではなく、文書を選ぶGVSM(Generalized Vector Space Model)[16]や、特異値分解法を用いて独立な要素のみを求めて次元数を減らしたベクトルを文書のインデックスとして用いる手法 [17] などの改良手法が提案されている。

さらに、検索精度を向上させるため、VSMにおける質問の修正方法として、利用者からのフィードバックを2回目以降の検索に反映させる手法が提案されている。この手法は、利用者とシステムとのインタラクションにおいて、利用者が適合すると判定した文書をシステムに教え、利用者の検索質問を修正する枠組みであり、適合性フィードバック (Relevance Feedback)[18]と呼ばれる。この手法の特徴は、

- 利用者からのフィードバック情報として文章を指定させることで、一度に多くの情報¹を検索要求に含めることが可能であること、
- 適合しない文書の指定を行うことで、負の情報をフィードバックできること、

にある。つまり、検索精度の向上のため、

- (4) 利用者との対話では1度に多くの情報が入力できること、
- (5) 適合しない負の事例が入力できること、

が重要であることを示唆している。

3 新聞記事検索システム

新聞記事検索タスクにおける対話の特徴を明らかにする目的で、まず、図1に示すような対話を想定し

¹文書を指定することで、利用者の知りたい情報を表わすキーワードが多く含まれていると考えられる。



「阪神優勝の記事をお願いします。」 「検索を希望される記事の概要をどうぞ。」
「吉田監督をお願いします。」 「該当する記事は全部で2038件あります。」
「14番をお願いします。」 「記事の範囲をしぼるため、追加するキーワードをどうぞ。」
「はい」 「該当する記事は全部で21件あります。」
「14番をお願いします。」 「記事の一覧を表示しても宜しいでしょうか？」
「はい」 「閲覧を御希望されるものの表題をどうぞ。」
「検索をご希望されたものと、一致しておりますでしょうか。」
「はい」

図 1: 対話例

て、対話データ収集システムを作成した。

システムからの「検索を希望される記事の概要をどうぞ。」というアナウンスに対し、利用者は「阪神優勝の記事をお願いします。」といったように、検索したい新聞記事の内容を音声で入力する。システムは入力された発声から検索キーワードを抽出し、新聞記事を検索する。その結果、検索記事数が多い場合はさらに利用者にキーワード入力を促すアナウンスを音声合成装置を用いて行い、検索記事数が適当な数であればタイトル一覧を表示し、その中から利用者に希望の記事を選択させる。この場合も音声入力を可能とする。最後に指定された記事内容を出力し、利用者の要求通りの記事か否かを確認する。

4 対話データ収集

対話データ収集で対象とした新聞記事は、1997年3月5日までの約10年間分の朝日新聞記事約180万件であり、検索エンジンは市販のものを用いた [19]。

4.1 収集方法

Wizard of OZ方式でデータの収集を行った。システムの代わりにするオペレータは、利用者の発声内容からキーワードを抽出して検索エンジンに入力する。次に、オペレータは検索結果に応じてアナウンス内容を決定し、音声合成装置を通じて利用者にアナウンスを行う。利用者への応答は、予め想定される表現内容

をテンプレートとして用意しておき、このテンプレートを
 選択することで行う。この際、パラメータ(例えば「該当する記事は全部でxx件あり...」の記事件数)がある場合は、その都度指定する。テンプレート数は、データ収集に先だてて行った試行実験の結果40種類とした。

また、データ収集前に概要を把握してもらうため、利用者には検索可能な新聞記事のタイトル例(12例)の提示した。しかし、提示記事に限らず自分の興味がある記事も検索可能である旨は説明した。

4.2 話者数および対話数

今回の実験で対話を収録した話者は男女計45人である。ほとんどの話者は実験に参加が容易な学生であり、平均年齢は約22歳であった。表1に示すように、収集対話数は245対話であり、収集した一人当たり平均対話数は約5.4である。

表1: 収集対話数および利用者数

話者数	45
対話数	245

4.3 発話数

利用者が1対話で発声した発話数の分布を図2に示す。これには、システムと利用者が最後に行う記事内容の確認のための発話も含まれており、全発話数は1019発話、1対話当たりの平均発話数は約4.2であった。図2より、1対話当たりの利用者の発話数は3回が最も多く、次に4回、5回の順であることが分かる。

4.4 キーワード数

発話表現の特徴を調べる目的で、1発話当り利用者が発声した検索キーワード数を調べた。検索キーワードを含んだ発話(利用者の全発話から、例えば「はい」や「お願いします」といったシステムからの質問に対するキーワードの含まれていない発声や、「3番目お願

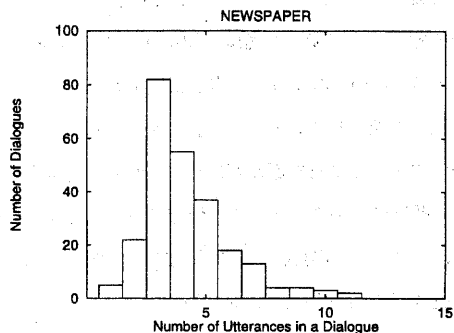


図2: 利用者の発話数分布

います。」といった検索結果一覧から目的の内容を選択する表現を除外した発話)に対し、形態素解析プログラム juman[20]を用いて形態素解析を行い、普通名詞、固有名詞、サ変名詞、未定義語、数詞、時相名詞(例:「今日」、「今週」、「将来」等)、名詞性名詞助数(例:「1000円」の「円」、「10日」の「日」等)をカウントした。なお、形態素解析では「Jリーグ」、「住専」、「野茂」、「ソ連邦」の4単語を追加した。その結果は、利用者が最初に発声した発話(以下、第1発話)のキーワード数は平均2.7個であり、それ以外の発話では2.2個であった。キーワード数の分布を図3に示す。

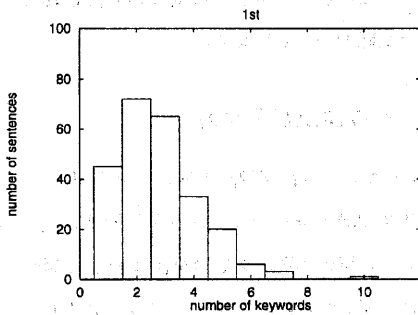
5 収集対話の分析

5.1 対話内容の分析

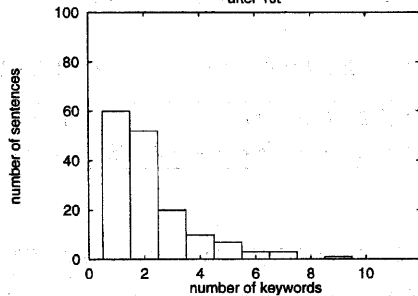
図2、図3より、収集した対話の特徴は以下のよう
 に考える。

- (1) 1回だけの発話で終わっている対話がある。
- (2) ほとんど対話は発話数が5回までであるが、中には長い対話もある。
- (3) 第1発話では、2および3個のキーワードが発話されるのに対し、第1発話以外では、1および2個が大半である。

以下には、上記の特徴に対して詳しく調べた結果を示す。



(a) 第1発話
after 1st



(b) 第1発話以外

図3: キーワード数の分布

5.1.1 発話数1回の対話

発話数が1回の対話は、検索した結果が空の場合であり、システムからは利用者に入力キーワードを変更させようと「申し訳ありませんが、該当する記事がありません。別の言葉をお願いします。」とアナウンスしたが、これに答えて利用者は別の内容の検索要求を行った対話であった。

例えば、

利用者: 「光ゲンジの解散」

システム: 「申し訳ありませんが、該当する記事がありません。別の言葉をお願いします。」

利用者: 「ドリームカムトゥルーの西川、事故を起こすをお願いします。」

5.1.2 発話数8回以上の対話

利用者とシステムとのやりとりの回数が多い例として、発話数が8回以上の対話(全13対話)を分析すると、以下のように分類できた。

- (1) 利用者が知りたい記事が複数あり、何度も記事の一覧表示から記事指定を行った。(6発話)
- (2) 検索した記事は利用者の知りたい内容ではなく、何度も記事の一覧表示から記事指定を行った。(3発話)
- (3) 検索した記事は利用者の知りたい内容ではなく、再度記事のしぼり込みのためのキーワード入力から繰り返した。(3発話)
- (4) 目的の記事をしぼり込むためのキーワード入力回数が多かった。(1発話)

5.1.3 キーワード数

図3に基づいて χ^2 乗検定を行った結果、第1発話と第2発話におけるキーワード数の分布の差は、有意水準1%で起こるべくして起こった(誤差の範囲ではない)と判定された。第1発話は「検索を希望される記事の概要をどうぞ。」とのアナウンスに利用者が答えたものであり、第1発話以外はほとんどが「該当する記事は全部でxx件あります。記事の範囲をしぼるため、追加するキーワードをどうぞ。」とのアナウンスに利用者が答えたものであり、いずれのアナウンスも利用者に自由にキーワード入力に行わせており、「xxx, yyy, zzzでお答えください」などの入力内容に制限を加えていない。しかし、第1発話以外ではキーワード数はほとんどが1つまたは2つと少ないことから、利用者は以前に入力した内容が正しく受け入れられ、次にわずかなキーワードを入力するだけで、目的の記事に到達できていると考えていることが伺える。

5.1.4 対話管理に関する考察

(1) 検索結果が空の場合

検索結果が空になった原因はほんとうに該当する記事が無かったのか、何らかの原因でたまたま記事が見つからなかったのかの判断がつかない。実際、利用者からの「ニューヨーク Yankees の優勝」という検索要求に対し、オペレータは「ニューヨーク Yankees」と聞き誤ったために、検索結果が空となった例があった。正しく「ニューヨーク Yankees」とキーワードを入力していれば、記事は発見できたはずである。このことは、現行のシステムからの応答内容では不十分であり、利用者への応答内容に工夫を行う必要があることを示している。

(2) 発話回数の多い場合

利用者とシステムのやり取りが多い対話は、利用者が見たい記事が複数ある場合と、該当記事までに到達するのに手間取った場合の2通りである。利用者が見たい記事が複数ある場合の対処は、今までの研究ではあまり考慮されておらず、新たにこのタスクにおける問題として捉えなければならない。具体的には、記事を何らかの基準でグルーピングを行い、そのグループ全体の記事を全て表示するなどの対処が必要だと考えられる。一方、該当記事までに到達するのに手間とった場合は、検索精度の問題であり検索精度を向上させる必要がある。そのため、第2章で述べたように適合性フィードバックなどの手法により、1回に利用者が指定できる情報(キーワード)をできるだけ多くできるように対話を制御する必要があると考えられる。

(3) キーワード数

利用者に何の制限も設けていないのに、第1発話以外で入力キーワード数が少ないことは検索精度に影響を与える。これは、システムとのやり取りの回数が多くなるだけでなく、音声認識誤りに対処しようとした場合、利用者から得られる情報が少ないことを意味する。そのため、音声入力による情報検索システムで

は、第1発話以外でも多くの情報が入力されることが重要な課題であると考えられる。

5.2 不要語及び言い直し

本タスクへの音声認識の適用の可能性を調べるため、音声認識の際に問題となる不要語及び言い直し・言い淀み表現の出現回数を全発話に対して調べた。その結果を表2に示す。表中の()内の数字は不要語の出現する発話の全発話に対する割合である。

表2: 不要語及び言い直し・言い淀みの出現回数

不要語を含む発話数	8(0.8%)
言い直し・言い淀みを含む発話数	1(0.1%)

表2より、全発話(1019発話)のうち1.0%程度の発話しか不要語が出現しないことが分かる。ここで、不要語は1発話中1回しか出現せず、出現箇所は全て文頭であった。今までの研究では、人と人との会話における不要語出現率(50%[21])、内線電話受付けシステムのような電話による人とシステムとの会話における不要語の出現率(7%[22])、オペレータ作業支援システムのような人と人との対話をモニターしながらボタン操作でシステムへ音声入力する場合の不要語出現率(約10%[23])などが報告されている。これらの結果に比べて、今回の収録した対話での不要語出現率は非常に低いと言える。

文献[22]では、不要語を含む発話はユーザがシステムに慣れるに従って減少したと報告されていることから、今回の収集データでは不要語の出現率が低い理由として、対話データを収録した利用者は20歳前後の若者であったため、情報検索システムには慣れておりシステムとの対話がスムーズに行えたためだと考えられる。実際、データ収録後のアンケートで約67%の人がパソコンをたまに使用するあるいは良く使用すると答えていた。しかし、このことについては年齢別にデータを収集する等さらに検討を要する。

6 まとめ

本稿では、新聞記事検索を対象としてシステムと利用者との対話を収集し、収集データの分析結果を示し、対話管理手法について考察を行った。対話管理方法には従来からの研究課題である、

- (1) 自然な対話のための応答表現の生成、
- (2) 検索結果が空であった場合の利用者への質問内容の選定、
- (3) 認識誤りへの対処、
- (4) 適合性フィードバックのような検索精度の向上のためのインタラクション方法、

の他に、今回の収集データから、

- (5) 第1発話以外では入力キーワード数が少ないことから、キーワード数を多くするための対話制御、
- (6) 利用者が見たい記事が複数ある場合の対処、

が必要であることが明らかになった。幸い、このようなシステムでは、豊富な情報をしかも即座に表示できるディスプレイが使用できることから、ディスプレイ上に表示した情報から利用者へ選択させる視覚的な対話方式も効果的であると考えられる。

また、収集データからは、不要語を含む発話は全発話に対して1%程度とほとんどないことが分かった。しかし、今回の実験では平均年齢約22歳と若い利用者が対象であったため、今後は年齢別に対話を収集して今回の実験結果と比較検討を行い、さらに詳しく本タスクにおける対話の特徴を明らかにしていく必要があると考えている。

本研究は通信・放送機構(TAO)からの受託研究「高齢者・障害者のための機能代行・支援システム技術の研究開発」の一環として実施した。

参考文献

- [1] J.Allen: "Recognizing intentions from natural language utterances", in Computational Models of Discourse, (Ed.

- M.Brady and R.C.Berwick), The MIT Press, pp.107-166, 1983.
- [2] S.J.Kaplan: "Cooperative responses from a portable natural language database query system", in Computational Models of Discourse, (Ed. M.Brady and R.C.Berwick), The MIT Press, pp.167-208, 1983.
- [3] B.J.Grosz and C.L.Sidner: "Attention, intentions, and the structure of discourse", Proc. of IJCAI85, pp.832-839, 1985.
- [4] 加藤恒昭, 中川慶: "質問応答における意図の把握と話題の管理" 情報処理学会, 自然言語処理研究会, 86-NL-58-6, 1986.
- [5] H.P.Grice: "Logic and conversation", in Syntax and Semantics III, (Ed. P.Cole and J.L.Morgan), Academic Press, pp.41-58, 1975.
- [6] 住田一男, 浮田輝彦: "質問応答システムにおける会話の自然性に関する考察" 信学論D, Vol.J70-D No.11, pp.2287-2293, 1987.
- [7] 近藤省造, 今村誠: "対話システムISDにおける対話モデルと協調的応答生成" 情報処理学会, 知能工学と人工知能研究会, 86-AI-57-3, 1988.
- [8] 西山敏雄, 大山芳史: "データベース検索における協調的な自然語対話処理と評価" 信学論D-II, Vol.J73-D-II No.4, pp.625-632, 1990.
- [9] 堂下修司: "音声・言語・概念の統合的処理による対話の理解と生成", 人工知能学会誌, Vol.12, No.1, pp.3-12, 1997.
- [10] 亀田弘之, 藤崎博也: "情報検索における音声・言語処理", 情報研報SLP16-11, 1997.
- [11] 中川聖一: "音声対話システム", 情報処理学会, 第52回全国大会, シンポジウム論文集, 1996.
- [12] 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, 音声言語情報処理研究会, 95-SLP-5-7, 1995.
- [13] 高橋敦子, 平井誠, 北橋忠宏: "自然言語によるデータベース検索における協調的応答生成", 信学技報NLC95-4, 1995.
- [14] 住田一男, 三池誠司: "知的情報検索の動向", 人工知能学会誌, Vol.11 No.1, pp.10-16, 1996.
- [15] G.Salton: "Automatic text processing", in The Transformation, Analysis, and Retrieval of Information by Computer, Addison-Wesley, Reading, Pennsylvania, 1989.
- [16] S.K.M.Wong, W.Ziarko and P.C.N.Wong: "Generalized vector space model in information retrieval", SIGIR Conference on Research and Development in Information Retrieval, SIGIR85, pp.18-25, 1985.
- [17] S.Deerwester, S.T.Dumais, T.K.Landauer, G.W.Furnas and R.A.Harshman: "Indexing by latent semantic analysis", Journal of the Society for Information Science, Vol.14, No.6, pp.391-407, 1990.
- [18] J.J.Rocchio, Jr.: "情報検索における適合フィードバック", Smart情報検索システム(監訳 神保健二), pp350-362, 1974.
- [19] 松下電気産業: "インターネット高速全文検索登録ソフトウェア PanaSearch/SSW マニュアル", 1996.
- [20] 松本裕治, 黒橋禎夫, 宇津呂武仁, 妙木裕, 長尾真: "日本語形態素解析システム JUMAN 使用説明書 version2.0", 1994-7.
- [21] 村上仁一, 嵯峨山茂樹: "自由発話音声認識における音響的および言語的な問題点の検討", 信学技報SP91-100, 1991.
- [22] S.Kuroiwa, K.Takeda, M.Naito, N.Inoue, S.Yamamoto: "Error Analysis of Field Trial Results of a Spoken Dialogue System for Telecommunications Applications", IEICE Trans. on Inf. & Syst., vol.E78-D, no.6, pp.636-641, 1995.
- [23] 井ノ上直己, 中村誠, 酒寄信一, 山本誠一, 谷戸文廣: "オペレータ作業支援システムのフィールド試験結果", 信学技報SP96-55, 1996.