

pLSA 学習を用いた音声認識誤りに頑健な質問応答選択手法

井上僚介
広島市立大学情報科学部

黒澤義明

目良和也
広島市立大学大学院情報科学研究科

竹澤寿幸

1. はじめに

音声対話システムとは、音声を用いて機械が人とコミュニケーションを図るためのシステムである。たとえば音声入力を受け付ける案内システムでは、ユーザはシステムに話しかけるだけでユーザが知りたい案内情報を得ることができる。このように、音声対話システムでは発話というごく簡単な操作で機械に指示や要求を与えることができる。この音声対話システムを実現するためには音声認識技術だけでなく、認識された音声に対して適切な応答を提示するための対話処理技術が欠かせない。

本稿では、小規模な質問応答データベースを用いた、音声による一問一答型質問応答システムに焦点を当てる。

小規模な質問応答データベースにおいては、ユーザがシステムに対して同じ意味で別の表現をすることがある。例えば天気を尋ねる表現として、「天気予報を教えてください」、「今日は晴れますか?」のようにいくつもの表現が考えられる。このような場合において、大規模な質問応答データベースならばデータベース上に入力と同様な質問例が存在するため、質問例を適切に選択することが容易である。しかし、小規模な質問応答データベースならば入力と同様な質問例が存在しないため、適切な質問例を選択できるとは限らない。更に、ユーザの発話音声は常に認識誤りのない正しい音声認識結果になるとは限らないため、誤りを含む入力から適切な質問例を選択する技術が必要となる。

そのための手法として、先行研究[1][2]における形態素マッチングを中心とした手法に対し、本研究では pLSA (Probabilistic Latent Semantic Analysis) 学習を用いることを提案する。pLSA 学習を行うことで、単語の出現頻度を元にして文章間の類似度を求めることができる。そのため、誤りを含む文章でも類似度を元にして適切な質問例を選択できるようになると期待できる。

また、音声認識した入力、質問例及び pLSA 学習を行う文書データを読み仮名表記に変換することも提案する。読み仮名表記に変換することで、認識した入力文が平仮名表記であり、質問例が漢字表記であるといった場合に、質問例を平仮名表記に変換することで、二個の文章が一致するといった判断をすることが可能になり、適切な質問例を選択できるようになると期待される。

本研究では、読み仮名変換を行った上で、先行研

究及び先述の pLSA を用いる手法を組み合わせることによって、実環境下での頑健な質問応答選択をさせることで、一問一答型質問応答システムの性能向上を高めることを目標とする。

2. 先行研究

2.1. 応答選択

用例ベースの応答選択では、質問例の応答データベースを用いて、入力に最も近い質問例から応答を生成する。具体的には、質問例と正解応答の対を QA ペアとしてデータベースに登録しておき、入力が生じたときにはデータベースから類似した質問例を選択することで、対応する正解応答をユーザに提示する。

質問例の選択方法として、先行研究[1][2]では、入力に対して最も類似した一例のみを質問例から参照し、適切な質問例を取得する、という簡便な最近傍法が用いられている。

2.2. 形態素マッチングによる類似度

文章間の類似度を求める方法として、文章を単語に分割した上で、一致した単語数を比較するという単語マッチングの手法が考えられる。日本語の文章においては、単語に分割することはすなわち形態素に分割することになるので、形態素マッチングとも呼ばれる。

先行研究では、複数文の集合、つまり入力と質問例の集合とを同時に比較して類似度を算出するため、平均単語数での単語マッチング[2]を考える。

この平均単語数での単語マッチングでは、一致する単語を数え上げたうえで文書の単語数で正規化を行う通常の単語マッチングと異なり、共通する単語の数、正規化係数に用いる単語数として文章一文あたりの平均値を用いるという点に違いがある。

この手法を用いて、平均単語数での単語マッチングによる文章 I 、 E の類似度 $s_w(I, E)$ を定式化することが出来る。

2.3. 認識結果質問応答データベース

実環境下での質問応答システムにおいては入力音声に対して誤りのない音声認識結果が常に得られるとは限らない。このときの入力にある一定の誤りの

傾向が存在すると仮定すると、この誤り傾向を質問例に含めることで、一定の誤り傾向に対して対応することが出来るようになるため、システムの応答の適切さは向上する、と期待される。

そこで、誤りを持つ入力に対する正解文書を用意し、それを QA ペアとしてデータベースを作成することで、認識結果質問応答データベースを作ることができる。

3. 読み仮名変換

通常、質問応答データベースの質問例や pLSA の学習に用いている文書などは漢字仮名混じり文で構成されている。しかし、このような漢字仮名混じり文には問題点がある。

例えば、質問例に「凄いね」という表現が登録されているとする。そこで、ユーザが同じように「凄いね」と音声で入力したとする。しかし、音声認識結果が「すごいね」となった場合、単純な形態素マッチングなどのような手法では適切な質問例を選べない場合がある。これを解決するため、入力や質問例をすべて読み仮名表記に変換することにした。読み仮名表記に変換した場合、質問例中の「凄いね」は「すごいね」となるため、入力と一致し、適切な質問例を選択することができるようになる。

また、後述の pLSA の学習に用いるコーパスに対しても読み仮名表記に変換した上で pLSA 学習に用いることとする。この処理を行うことで、同音同義語であるにも関わらず、漢字表記や読み仮名表記で統一されていない場合でも、読み仮名表記に統一することで、文の類似性を正しく取得できると期待される。

4. pLSA による学習

本研究では、認識された入力と質問例との文章間類似度を算出し、その類似性から選択する質問例を確定させるために pLSA 学習を行わせる。pLSA 学習を行うことで、入力と質問例が一致していない場合でも、類似性から適切に質問例を選択できることが期待される。

4.1. LSA

LSA とは、潜在的意味解析 (Latent Semantic Analysis) のことで、ベクトル空間における主成分分析を利用した情報検索の手法である[3][4][5]。

共起行列と呼ばれる単語文書行列を与え、その行列の特異値分解を行い、いくつかの主成分を除き新たに行列を再構成する。このようにして、元の文書の内容をより低次元のベクトル空間上の座標として

表現する。つまり、LSA は表現するベクトルを圧縮することでデータ中のノイズを取り除くことになる。

これを利用することで、同じ意味を持つ固まりの抽出(類義語の抽出)及び、各々の単語がどの意味に関係しているかなどを算出することが出来る。

しかし LSA には問題点がある。具体的には、共起行列の適切な意味的關係を抽出するには、単語に対して重み付けをしておく必要があるということである。

4.2. pLSA

LSA に対して、pLSA とは、確率的潜在意味解析 (Probabilistic Latent Semantic Analysis) のことで、基本的には LSA と同様に次元の圧縮を行うだけでなく、次元圧縮を確率的に行う手法である[3][4][5]。

潜在変数 $z \in Z$ を考えて、文書 d における単語 w の生起確率は、

$$P(d, w) = \sum_{z \in Z} P(z)P(d|z)P(w|z)$$

と表せる。

また、潜在変数モデルにおける最尤推定のために、EM アルゴリズムを利用することで以下のように定式化できる。

まず、E ステップとして次式が定式化される。

$$P(z|d, w) = \frac{P(z)P(d|z)P(w|z)}{\sum_{z' \in Z} P(z')P(d|z')P(w|z')}$$

次に M ステップとして以下が定式化される。

$$P(w|z) \propto \sum_{d \in D} n(d, w)P(z|d, w)$$

$$P(d|z) \propto \sum_{w \in W} n(d, w)P(z|d, w)$$

$$P(z) \propto \sum_{d \in D} \sum_{w \in W} n(d, w)P(z|d, w)$$

ここで、 $n(d, w)$ は文書 d における単語 w の出現回数とする。

pLSA ではこの E ステップと M ステップを反復させることで、生起確率 $P(d, w)$ を最大化させるようなモデルが作成されることになる。また、E ステップの右辺全体を β 乗するような湿度パラメータ β ($0 < \beta \leq 1.0$) を与えると、 $\beta = 1.0$ に近ければ近いほど、生成される確率モデルの確率分布は鋭いピークを持つようになり、逆にこの値を小さくすると、平滑化される。

4.3. pLSA による学習の実装

pLSA を用いた学習をさせるために、pLSA の計算ツール[6]を用いた。このツールを用いることで、ある文書と別のある文書における共起確率を求めることができる。

例えば、「今日の天気はどうか」という文書があるとする。これを形態素解析すると、「今日の天気はどうか」と分割される。この単語それぞれにインデックスを割り当て、単語の出現頻度を記した共起行列を作成する。

この共起行列、つまり文書ベクトルを作成した上で、先述の pLSA の計算ツールに用いると、学習データが生成される。この時に生成される学習データを今後の計算に用いる。

このとき、共起行列を作成するために用いる文書として、質問応答データベース中の質問例、音声認識された入力、及び学習用のコーパスとして新聞記事[7]を用いる。新聞記事については、学習対象を一日分、一ヶ月分など小分けして用いる。また、共起行列を作成する際には、その行に出現する単語の頻度が必要となるので、新聞記事を、一行一記事、一段落、一文に分けた文書を用いる。

5. 実験結果

5.1. 実験

小規模な認識結果質問応答データベースを用いたシステム応答の適切さを評価する実験を行った。

使用する音声データとして、本学の 2009 年度オープンキャンパスの際に本システムのプロトタイプを設置して収集した音声データを使用する。このプロトタイプシステムは広島市と広島市立大学および対話エージェントに関する情報のやり取りが可能であり、146 個の質問例を有する小規模な質問応答データベースを使用している。

この際に収集した音声データ 355 件を用いてクロードテストを行う。また、音声認識エンジンには Julius[8][9]を用いる。

応答正解率の算出方法として、その音声に対する正解文書を用意しておき、その正解文書とシステムが選択した質問例を比較し、正解かどうかを判断する。

5.2. 実験結果

まずは pLSA で学習させる文書量を決定する。pLSA 学習には新聞記事を用いるため、一日分、一週間分、一ヶ月分で実験を行う。その結果を示す。(表 1)

表 1 : 新聞記事の量による比較

	一日	一週間	一ヶ月
総数	355	355	355
正解数	150	125	92
正解率	42.25	35.21	25.92

表 1 より、一日分の新聞記事のデータで十分有用な結果が得られることが分かる。これは、質問応答データベースが小規模であるのに対し、新聞記事のデータが大規模であるため、データ量が多くなるにつれ、質問例や入力に存在する特徴が大量の新聞記事に吸収されてしまい、質問例の特徴が打ち消されてしまったためと考えられる。この結果から、よりデータ量の多い、一年分の新聞記事データで同じ処理を行った場合、質問例の特徴が打ち消されてしまい、良い結果は得られないと判断したため、一年分のデータでの実験は行っていない。

また、今回は pLSA の学習のために新聞記事を用いているため、処理対象の日によって記事の内容は変わってくるため、学習結果に差異が生じるのではないかと考えられる。そこで、ランダムに一日分選択し学習させ、結果に差が現れるかどうかを実験により確認する。(表 2)

表 2 : 一日分の記事をランダムに選択し、比較

	A	B	C
総数	355	355	355
正解数	152	156	154
正解率	42.82	43.94	43.38

※A、B、Cそれぞれ学習対象の記事をランダムに選択

この結果から、結果はまったく同じにならないものの、ほとんど差異が見られないことが分かる。これは、学習対象の記事の量、記事中に存在する語彙の種類ともに大きな差がないために、pLSA で類似度を求める際、単語の出現頻度に大きな差がないために、このような結果になったと考えられる。

次に、pLSA 学習に用いる共起行列を作成する際に、どのような処理単位を一単位として扱うと良いかを検討する。そこで、学習対象である新聞記事の処理単位を一記事、一段落、一文の計三通りに対して実験を行った。(表 3)

表 3 : 新聞記事の切り分け方による比較

	記事	段落	文
総数	355	355	355
正解数	172	150	162
正解率	48.45	42.25	45.63

表 3 から、一記事一単位として共起行列を作成するほうが有用であると分かった。これは、仮に一単位を一文とした場合、一文が短すぎて、類似する単語を抽出することがうまく出来ず、効率的な次元縮約が出来なかったためではないかと考えられる。逆

に、一個の記事を一単位とすると、一単位が長い
ため、類似するような単語を多く抽出でき pLSA で効
率的に次元縮約できるからではないかと考えられる。

以上の結果から、pLSA 学習に用いる新聞記事デ
ータは、「一日分」かつ、「一個の記事を一単位」と
して共起行列を作成することが最も良いと考えられ
る。

これらを元に性能評価実験を行う。先行研究の形
態素マッチングをベースとした質問例選択手法 (①)、
新聞データを含めた pLSA 学習による質問例の選択
手法 (②)、先行研究と pLSA 学習を組み合わせる手
法 (③)、読み仮名変換を行った上で先行研究と pLSA
学習を組み合わせる提案手法 (④)、の計 4 通り対
して実験を行う。この実験の結果を示す。(表 4)

表 4 : 性能評価実験

	①	②	③	④
総数	355	355	355	355
正解数	168	169	171	180
正解率	47.32	47.61	47.89	50.70

- ① 先行研究の質問例選択手法
- ② pLSA 学習による質問例選択手法
- ③ 先行研究+pLSA 学習による質問例選択手法
- ④ 読み仮名変換を加えた③の手法 (提案手法)

表 4 の結果から、提案手法は先行研究に比べ、お
よそ 3 ポイント程度の性能向上を確認することがで
きた。良くなった例として、音声認識結果が「三日
は」で、本来は「こんにちは」と発話したケースな
どが、提案手法では適切に質問例を選択することが
出来ていた。これは、「にち」という部分に類似する
と判断したためと考えられる。また、読み仮名変換
の際に期待されたように、認識結果が平仮名で質問
例中には漢字仮名混じり文であった場合など、適切
に選択できた。しかし結果が悪くなったケースとし
て、質問例中に意味は異なるが文として一文字しか
差異が存在しない質問例が存在する場合、例えば「今
何時ですか」「今何歳ですか」という質問例が存在
する場合などは、提案手法では適切に質問例を選択
することが出来なかった。

また、極端に短い音声の場合、入力音声
が認識誤りなく正しく認識されているにも関わらず、
適切な質問例を選択することが出来ない事例が数多く
みられた。このような場合に対しても解決が必要である。

6. まとめと今後の課題

本研究では、先行研究の形態素マッチングをベ
ースとした質問例の選択手法に加え、pLSA による学
習及び読み仮名変換を用いた質問例の選択手法を提

案した。

実験結果から、先行研究の形態素マッチングをベ
ースとした手法に比べて、先行研究の手法に加え、
pLSA による学習、及び読み仮名変換を組み合わせ
た提案手法の方がおよそ 3 ポイント程度の性能向上
を確認できた。

具体的には、音声認識結果が平仮名となってい
ない、質問例と一致しなかったケースが適切に質問
例を選択できるようになっていた。これは、質問例な
どの読み仮名変換処理の効果と考えられる。また、
先行研究の手法では、単語が一個一致しただけでは
適切に質問例を選択することが出来なかったが、提
案手法では適切に質問例を選択することが出来てい
た。これは、文中に類似する単語が出現している、
と判断されたために、類似する質問例と認識し、適
切に選択することが出来たと考えられる。

また、本研究では pLSA 学習のための言語資源と
して、一日分の新聞記事を用いたが、新聞記事は話
し言葉ではなく書き言葉に最適化されている。そこ
で、話し言葉に最適化されている文書を用いて同様
な評価実験を行うことが考えられる。例えば人間の
会話を文書化して学習に利用することによって、よ
り話し言葉に即した最適な結果が得られるのでは
ないかと期待される。よって今後は pLSA の学習に用
いる文書データを検討し、性能向上を図る。

参考文献

- [1] 竹内翔大, ツィンツアレク トビアス, 川波弘道,
猿渡洋, 鹿野清宏. 音声案内システムにおける質問
応答データベースの最適化. 情報処理学会研究報告,
2007-SLP-69(51), 2007.
- [2] 竹内翔大. 音声対話システムにおける
Leave-one-out 交差検定法を用いた質問応答デー
タベースの最適化手法. 奈良先端科学技術大学院大学
修士論文, NAIST-IS-MT051098, 2008.
- [3] Thomas Hofmann. Probabilistic Latent
Semantic Analysis. Uncertainty in Artificial
Intelligence, 1999.
- [4] 栗山直人, 鈴木基之, 伊藤彰則, 牧野正三. PLSA
言語モデルの学習最適化と語彙分割に関する検討.
情報処理学会研究報告, 2006-SLP-60(8), 2006.
- [5] 持橋大地, 松本祐治. 意味の確率的表現. 情報処
理学会自然言語処理研究会, 147-12, pp. 77-84, 2002.
- [6] <http://chasen.org/~taku/software/plsi/> .
- [7] 毎日新聞社. 毎日新聞地方版. 2005.
- [8] 河原達也, 李晃伸. 連続音声認識エンジン Julius.
人工知能学会誌, Vol. 20, No. 1, pp. 41-49, 2005.
- [9] 李晃伸, 河原達也. Julius を用いた音声認識イ
ンターフェースの作成. ヒューマンインターフェ
ース学会誌, Vol. 11, No. 1, pp. 31-38, 2009.