

# 質問応答対の増加による精度低下及び速度低下を軽減する選択手法

井上僚介

黒澤義明

目良和也

竹澤寿幸

広島市立大学大学院情報科学研究科

[inoue, kurosawa, mera, takezawa]@ls.info.hiroshima-cu.ac.jp

## 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、すなわちユーザの要求に対して適切な応答を返すための対話処理技術並びに大量の質問応答データベースが必要となる[1]。

たとえば、この質問応答データベース (QADB) に対して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。たと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。このような時、質問応答データベースが小規模である場合には、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。この問題を解決させるために、pLSA (Probabilistic Latent Semantic Analysis) 学習を用いる手法を提案した[2]。しかしこの手法は pLSA 学習での次元縮約に左右されやすいという問題点が指摘されている。

そこでこの問題点の解決のために、次元縮約における次元数が質問例選択の精度にどのような影響を与えるか、また、計算時間はどのように変化するかを予備的に調査した。その結果、pLSA 学習の次元数を増加させると精度向上が見られる。反面、計算時間の増大に繋がることが判明した。応答までにかかる時間が増大するという事は、一問一答型の質問応答システムでは致命的なことであり、また、pLSA 学習による質問例選択手法の限界であるといえる。

そこで本稿では、実環境下における質問例選択の頑健な手法、すなわち、質問応答対の増加に対して精度低下及び速度低下を軽減する手法を提案する。具体的には、拡張された質問応答データベースに対しクラスタリングを行い、入力の属するクラスタを決定する。そして入力の属するクラスタのみに限定して形態素マッチング手法を適用することで計算速度の向上を狙いつつ適切に質問例を選択できる。この手法により実環境下での質問応答システムにおける精度低下及び速度低下を軽減させることを目指す。

## 2. 先行研究

本研究に関連する研究及び先行研究について紹介する。一点は一問一答型質問応答データベースにおける質問例選択の基本となる形態素マッチングによる手法、もう一点は、大規模な言語資源を用いた pLSA 学習による選択手法である。

### 2.1. 形態素マッチングによる質問例選択手法

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、入力が生じると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。

質問例の選択手法として、先行研究[1]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。この際、文章間の類似度を求める手法として、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と同義となるので、これは形態素マッチングとも呼ばれる。この手法を拡張し、単語数の平均をとり、更に音声入力にはある一定の誤り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可能になる。これらの手法を用いてシステム応答の適切さを向上させる。

### 2.2. pLSA 学習を用いた応答選択手法

先述の形態素マッチングを中心とした手法に対し、pLSA 学習を用いる質問例選択手法[2]が挙げられる。先述の手法では一定の音声入力の誤りには対応できるが、小規模な質問応答データベースの元では、質問者の多彩な発話表現に対応できない問題点があった。そこで、小規模な質問応答データベースであるという条件下でも質問例選択をより適切に行うために、pLSA 学習を用いた手法が提案された。この手法は、新聞記事という大規模な言語資源を小規模な質問応答データベースとともに学習させることにより、類似する表現を同一表現として取り扱うことによって質問例選択精度を向上させるという戦略をとっている。これにより従来の形態素マッチングを中心とした手法と比較して精度向上が可能となる。

### 2.3. pLSA

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].

潜在変数  $z \in Z$  を考えて、文書  $d$  における単語  $w$  の生起確率は以下のように表せる.

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

また、潜在変数モデルにおける最尤推定のために、EM アルゴリズムにより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 E ステップと M ステップを反復させ、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. また、E ステップの右辺全体を  $\beta$  乗するような温度パラメータ  $\beta$  ( $0 < \beta \leq 1.0$ ) 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.

### 2.4. 従来手法の欠点

pLSA 学習を用いた応答選択手法では、学習の際に次元縮約を行う. この際、N 次元に縮約する必要があるがこの縮約次元数の最適なパラメータが確定していないという問題点が残されている. そこで、次元数パラメータを変更して実験を行った. 結果、次元数を増加させれば増加させるほど精度が向上することが判明した. 反面、計算時間が増大することとなった. 一般的に、一問一答型質問応答システムにおいては質問に対して素早く応答を返すことが非常に重要である. そのため、計算時間の増加は致命的であるといえる.

## 3. 提案手法

前節より、pLSA 学習を用いた場合、質問応答データベースの拡張に加えて更なる精度向上を求めた場合、計算時間の増加を招くという問題点が生じた. この問題は質問応答データベースの増加による質問例の多彩化により顕著になると考えられる.

そこで、本研究では次のような手続きにより、問題点の解消を試みる.

1. 質問例のクラスタリング
2. マッチング手法

質問応答データベースを拡張により、更に広範な意図の質問に対して適切に応答を返せるようになる. しかし、データベースを拡張することにより、既存手法では精度低下及び計算時間の増加の懸念がある. そこでこの手続きを適応することにより問題点の解決を図る. 具体的には、各質問例の属しているクラスタを取得し、入力属するクラスタに限定してマッチング処理を行う. クラスタリング処理により質問例の属するクラスタを確定し、後のマッチング処理における計算時間の減少が期待できる. 更に、マッチング処理を加えることにより質問例を適切に選択できると期待される.

ここで、以下のような入力例並びに質問例が与えられていたとする.

入力

Input. 今日の天気予報は

質問例

1. 天気は
2. 空模様は
3. 今日の空模様は
4. 今週の天気予報は

この例について各手法について検討する. 形態素マッチングを基本とした手法では、入力に対し、形態素マッチング処理により候補が2つ挙げられることになる. そのため、類似度が同一となってしまうことで適切に正解となる質問例を選択できない可能性がある. また、新聞記事による pLSA 学習を用いた手法の場合、この例における質問例全てが同じ意味の質問として縮約されてしまう可能性がある. ゆえに、適切に質問例を選択できない可能性がある. これに対して、提案手法ならば、まず、質問例と入力を合わせてクラスタリングを行う. すると、[1, 4, Input], [2, 3]というクラスタに分けることができると考えられる. そして入力が属しているクラスタ内で形態素解析を行った上で形態素の一致数元に類似度を算出するという形態素マッチング手法により質問例選択を行う. これにより、精度の向上だけでなく、計算時間の減少すなわち応答速度の向上が期待される.

それぞれの手続きを次に述べる.

#### 3.1. 質問応答データベースの拡張

本研究では初めに、既存のデータベース (既存 QADB) に新たなデータベース (追加 QADB) を追

加した拡張データベース（拡張 QADB）の作成を行うこととした。（表 1）これにより、より多種多様な質問例に対応可能になる。なお、今回追加した質問応答データベースは NICT 提供の京都の観光案内対話コーパス[4]より質問に対して応答を返しているような会話のセットを抽出し、質問応答対を作成することで質問応答データベースを作成した。この質問応答データベースは京都の観光情報が主となっているため、広島県の観光情報に書き換えた上で作成した。

表 1：質問応答データベースの拡張

	既存 QADB	追加 QADB	拡張 QADB
件数	141	70	211

### 3.2. 質問例のクラスタリング

本研究では質問例クラスタリングには先述の pLSA 学習による選択手法と同様に、pLSA を用いてクラスタリングすることとした。また、pLSA の問題点である次元縮約における次元数は 70 とした。これは、本研究における既存の質問応答データベースに加え新規に追加する既存のデータベースとは異なるデータベースの質問例数が 70 件であるため、既存のデータベースで一つのクラスタ、そして追加した異なるデータベースが各々一つのクラスタを形成することを期待している。この質問例クラスタリングの際にユーザから渡された音声入力結果もまとめてクラスタリングを行うことで、入力の属するクラスタを検出、クラスタが決定した場合に質問例選択手法を適用することとした。pLSA を用いたクラスタリングでは先述の温度パラメータ  $\beta$  の設定によりクラスタリングの性質を変更することが可能である。例えば、温度パラメータ  $\beta=0$  に近付けると分布が平滑化され、ソフトクラスタリングに似た結果が得られる。また、 $\beta=1.0$  に近付けると分布がより鋭くなり、ハードクラスタリングに似た結果が得られる。今回は曖昧なクラスタリング結果では質問例選択において入力の属するクラスタを決定させることが難しい。入力が属する可能性のあるクラスタ全てにおいて計算等を行うと計算時間の増大を招く。よって曖昧なクラスタリングは不利であると考えたため、よりハードクラスタリングへ近付けるため、 $\beta=1.0$  とした。なお、pLSA の計算等には提供されている計算ツール[5]を用いた。

具体的には、拡張された質問応答データベースの質問例と入力をまとめて pLSA による計算を行う。その前段階として形態素解析を行い各質問例に対して単語の出現回数を記した単語文書行列を作成する。この際、形態素解析は品詞による区別を行わずに単語文書行列を作成した。これは、今回利用する品詞を動詞や形容詞といった品詞に限定した場合に、短い質問例では単語が 1 つだけとなってしまう、その

ような質問例が増加した場合に pLSA ではクラスタリング結果に影響を受けやすいと考えた。品詞に関わらず単語文書行列を作成し pLSA により計算する。これにより、pLSA を用いて質問例及び入力をクラスタリングすることで、入力の属するクラスタを抽出することが可能となる。

### 3.3. マッチング手法

クラスタリングのみでは適切な質問例を選択できるとは限らないため、クラスタリングを行った上で形態素マッチングによる質問例選択を行う。入力と同じクラスタに属する質問例に対してのみ行うことで、形態素マッチングを行う必要のある質問例の総数が限定されるため、計算量の削減が見込める上、誤選択率も低下すると期待される。なお、本研究では形態素マッチングにおいて pLSA によるクラスタリング時の計算同様、全ての品詞を利用した。また、形態素マッチング手法に関しては先行研究と同様な手法を用いることで、pLSA を用いたクラスタリングがどの程度有効かを検討するための材料とした。

### 3.4. 提案手法と先行研究の比較

提案手法の利点として、大量の言語資源を必要としないことにより計算時間の短縮が見込める点、また、既存の質問応答データベースに対して新たに質問応答データベースを追加しても初めにクラスタリングを行うため質問例選択にある程度効果的であるという点が挙げられる。しかしクラスタリングの結果に左右されやすく、クラスタの誤分類の懸念が残る。この点については、実験を行って検証する。

## 4. 実験結果

### 4.1. 実験

既存の質問応答データベースに新たに異なる質問応答対を追加した質問応答データベースに対して、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及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る質問応答データベースを使用している。これを既存の質問応答データベースとし、それに広島市に関する観光情報等、既存のデータベースとは質問の性質が異なる質問応答対を追加し実験を行う。

評価に用いる音声データは先述の 2009 年度本学のオープンキャンパスで収集された音声データとし、その件数は 355 件である。音声認識エンジンには Julius[6]を用いてクロズドテストを行った。

応答正解率の算出方法には、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。

## 4.2. 実験結果

本実験では同じ条件下でいくつかの手法について実験を行う。形態素マッチングを基本とした先行研究の手法 (①), pLSA 学習を用いた先行研究の手法 (②, ③). ①と②については既存の質問応答データベースでのみ実験を行い, ③については既存の質問応答データベースを拡張した, 拡張質問応答データベースを対象とした。また, 本研究における提案手法を④とした。これらの実験結果を示す。(表 2)

表 2 : 性能評価実験

	①	②	③	④
正解数	168	180	151	199
正解率	47.3	50.7	42.5	56.1
時間(sec)	4.99	9.87	10.1	3.87

表 2 から, 質問応答データベースを拡張することにより, 既存の手法では精度低下の危険性が生じることが分かる。また, 拡張された質問応答データベースに対し, 本研究で提案した手法を用いると精度が向上していることが分かる。

また, 1 つの入力データに対して応答を返すまでの計算時間<sup>1</sup>に関しては②や③に比べて④の手法では計算時間が短くなることが分かった。計算時間の点でも提案手法が有効性が確認された。特に, ①と④を比較すると, マッチング手法が同じであるにも関わらずクラスタリングにより精度向上及び計算時間の短縮が達成されていることから, 提案手法の有効性が確認された。

## 5. まとめと今後の課題

本研究ではこれまでの小規模な質問応答データベースに対し, 新たに質問の性質が異なる質問応答対を追加した拡張された質問応答データベースに対し, 精度低下並びに計算時間の増大を軽減するための手法の検討を行った。質問例のクラスタリングの後に質問例選択手法を適用することにより, これまでの手法のように精度低下や計算時間の増大を軽減することが実験により確認された。特に大規模な言語資源を必要としない提案手法は計算時間の削減に大いに貢献しており, 提案手法の有効性が確認された。

特に, 先行研究で用いられた, pLSA 学習による質問例選択手法では, 質問応答データベースを拡張すると精度が低下していることが確認された。これは, 追加した質問応答対に単語数が長い質問例が存在することによって, 新聞記事を利用した学習に影

響を与え, また, 形態素解析でも類似すると判断されたことが原因ではないかと考えられる。

また, 精度が全体的に低い原因として, 本実験で実験に用いている音声データが不十分なことが挙げられる。音声データを音声認識すると音声認識結果が適切でない場合が多く, 新たに音声データの収集が必要であると考えられる。また, 音響モデル等に改良を加えることでも更なる性能向上が期待される。更に, アルゴリズム等の改良を行う事によって計算時間の短縮も期待される。

しかし提案手法の問題点として, 拡張された質問応答データベースに対してクラスタリングを行う際, 今回は pLSA を用いたが, pLSA はパラメータによって結果に差が生じ, また, 表層単語のみを元にクラスタリングを行っているため, 正確なクラスタリングであるとは限らない。故に, より正確なクラスタリングが行える手法の検討が必要であるといえる。

また, 今回既存の質問応答データベースに新たに観光に関する対話から抽出した質問応答対を追加することで拡張したデータベースを用いて実験を行ったが, この際に追加しているデータベースがどのようなデータベースでも本研究での報告のように良い精度が得られるのかどうかは疑問が残る。例えば, 観光に関する対話ではなく病院等で行われているような対話から抽出した質問応答対を用いた場合に同様な結果が得られるのだろうか。そのようなデータを追加して拡張したデータベースに対して同様な結果が得られるかどうか, 今後の課題とする。

今後は実験のためにより多くの音声データの手法とともにクラスタリング手法の検討等を行い, より実用的な一問一答型質問応答システムの構築を行う。

## 参考文献

- [1] 竹内翔大, ツインツアレク トビアス, 川波弘道, 猿渡洋, 鹿野清宏. "音声案内対話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] 井上僚介, 黒澤義明, 目良和也, 竹澤寿幸. "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," 言語処理学会年次大会, 2010.
- [3] Thomas Hofmann. "Probabilistic Latent Semantic Analysis," Uncertainty in Artificial Intelligence, 1999.
- [4] 大竹清敬, 堀智織, 柏岡秀紀, 中村哲. "京都観光案内対話コーパスにおける対話行為の分析," 言語処理学会年次大会, 2008.
- [5] <http://chasen.org/~taku/software/plsi/> .
- [6] 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.
- [7] <http://mecab.sourceforge.net/>

<sup>1</sup> 形態素解析器 MeCab[7]の計算時間含む