論　文

tf・AoI の重み付けに基づく類似性を用いた話し言葉における質問文の同定法

木村　泰知[†a)]　　荒木　健治[††]　　栃内　香次[†††]

Identification of Spoken Question Using Similarity Based on $tf \cdot AoI$

Yasutomo KIMURA[†a)], Kenji ARAKI[††], and Koji TOCHINAI[†††]

あらまし　類似度は情報検索や情報抽出で利用されているが，対話処理にも適用可能である．音声対話処理では音声認識誤り，間投詞，雑音が含まれ，同一の表現が行われることはまれである．そこで，これらの現象を考慮しながら，入力文と類似している文を見つける必要がある．本論文では $tf \cdot AoI$（term frequency × Amount of Information）の重み付けに基づく類似性を用いた質問文の同定法を提案する．入力文に含まれる単語に対して，単語頻度 × 情報量の重み付けを行い，ユークリッド距離の計算結果から，最も類似性の高い質問に対応している応答を出力する．比較実験により，“入力文との一致率による比較” に対して 13 ポイント，“$tf \cdot idf$ の重み付けによる類似度” に対して 6.5 ポイントの向上が確認された．

キーワード　$tf \cdot AoI$，類似性，質問文の同定，情報量，ユークリッド距離

1. まえがき

近年，情報抽出，情報検索，要約の研究が注目されている [1]．これらの研究では精度向上のために情報の分類が行われ，分類には類似度が利用されている．類似度は分類問題の尺度だけでなく，文書の比較にも用いられ，応用システムである対話処理にも利用できる．このような背景から，類似度の測定は応用分野において必要不可欠な技術と認識されている．

類似度は，文あるいは単語の比較を行うための基準として利用される．そのため，ある 2 文の類似度を考えた場合，完全な一致（Exact match）が最も高い類似度となる．しかしながら，完全に一致していない場合，一致率の高い文が必ずしも類似しているとは限らないので，それぞれの対象に適した表現方法あるいは，計算方法が選択される．検索において類似度を測定する手法は多く提案されており，ブーリアンモデル，ベクトル空間モデルなどの検索モデルがある [3]．ブーリアンモデルは検索質問を論理式で表現し，ブール代数を基礎としているが，基本的には入力文がそのままの形で文書中に出現することを要求するため，一致率による比較と同じことになる．ベクトル空間モデルでは索引語の出現頻度を利用し，ユークリッド距離，余弦，Dice 係数などを用いて類似度が計算される．更に，特徴を反映させるために出現頻度と他の重みを掛け合わせる $tf \cdot idf$ [2], [3] と同様の重み付けが行われている．

ところで，多様な表現が含まれる対話処理では，頑健性を必要とする．対話処理において，頑健性を高めるために古くからキーワードが利用されている [4]．特に，音声対話処理では音声認識誤り，間投詞，雑音が含まれるため，キーワードが利用される傾向が強い [5]〜[7]．ハンバーガーショップの音声対話を想定した場合，「えーと，ハンバーガー とー，うーんと ポテト ください」では，下線部分の単語がキーワードとなる [8]．ここでキーワードとは発話中の重要な単語，句であり，それ以外は不要語とされる．この場合，新商品や季節限定商品に応じてキーワードの変更・追加・

† 北海道大学大学院工学研究科，札幌市
Graduate School of Engineering, Hokkaido University, Kita 13 Nishi 8, Kita-ku, Sapporo-shi, 060–8628 Japan
†† 北海道大学大学院情報科学研究科，札幌市
Graduate School of Information Science and Technology, Hokkaido University, Kita 14 Nishi 9, Kita-ku, Sapporo-shi, 060–0814 Japan
††† 北海学園大学大学院経営学研究科，札幌市
Graduate School of Business Administration, Hokkai-Gakuen University, 4–1–40 Asahi-machi, Toyohira-ku, Sapporo-shi, 062–8625 Japan
a) E-mail: kimu@media.eng.hokudai.ac.jp

削除を人手で行う必要がある．英語では冠詞，日本語では助詞などが不要語とされているが，長い文の場合，多くの不要語（助詞等）を無視することになる．

そこで，本論文では従来不要語と呼ばれていた単語も考慮しながら，対話処理の一つである質問応答について述べる．質問応答の場合，一般に，質問応答のデータを利用して，入力された質問文に適した応答が行われる [9]．同じ内容の質問に対して様々な表現が存在するため，異なる表現で同一内容の文を判断する必要がある．これらの理由から，本論文では情報量の重み付けを利用した類似度による同定を行う．情報量の式は $-\log_2 P(x) = -\log_2 \frac{f(x)}{N}$ であり，N は総単語数，$f(x)$ は x がデータ中に出現する頻度である．本手法では，情報量による重み付けを用いる．ここでは，情報量を利用しているため，出現頻度が低いほど大きな重みとなる．低頻度語の問題として，訳語の自動抽出における研究 [10] では，茶筅 [11] により名詞と判定した単語で頻度 1 の語が，全異なり単語の半数近くを占め，低頻度語を利用した自動対訳抽出の問題点として挙げられている．このような同じ出現頻度で，同じ出現アラインメントの単語が存在した場合，訳語抽出が困難となる．他にも，“to be or not to be” のように高頻度語から構成されたフレーズが特別な意味をもつこともある [3]．このように，低頻度語だけでは処理できない単語も存在するが，単語分割以外は言語知識を利用しないため，未知語が多く存在する場面にも処理が容易で，汎用性が高いといえる．頻度を利用した重み付けは，古くから利用されており，高頻度語は機能語，低頻度語は内容語として利用されている．このような特徴を生かして，低頻度語を「手掛り」とした手法の有効性が確認されている [12]．本手法は，情報量による重みと単語の出現頻度を掛け合わせ，これを $tf \cdot AoI$（$termfrequency \times AmountofInformation$）と呼ぶ．

本手法では，ベクトル空間モデルの考え方を利用している．ベクトル空間モデルにおいて，特徴量の設定は結果に大きな影響を与える．従来のベクトル空間モデルでは，不要語処理を行い，選択された単語を特徴量の属性としていた．しかし，不要語と呼ばれる語が必要とされる場面も少なくないため，我々は入力文中の全単語を考慮しながら，2 文の類似性を測定する，$tf \cdot AoI$ の重み付けを用いた測定手法を提案する．本論文では，類似度を用いて質問文の同定を行う．質問応答や FAQ のような対話処理を考えた場合，質問文の同定として，類似度が利用されており，ダイアログナビ [9] や効果的マッチング法 [13] がある．ダイアログナビの類似度は，書き言葉を対象としており，係り受けの被覆率を利用しているため，話し言葉における類似度の効果が少ないと考えられる．効果的マッチング法 [13] については，質問文から応答文を導くために $tf \cdot idf$ を利用して順位付けを行っている．

情報検索の重み付けにおいて，$tf \cdot idf$（term frequency × inverse document frequency）[2], [3], [13] は広く利用されており，キーワードが文書中に含まれる頻度とキーワードを含む文書数の逆数の対数をとり，掛け合わせた結果を重みとしている．$tf \cdot idf$ は，他の文書に出現しないタームほど，文章の特徴を表すキーワードとして重みを大きくする．しかしながら，$tf \cdot idf$ の重み付けでは，少ない単語から構成される文章のように idf に差が生じにくい場合には，適切な重み付けを行えない欠点をもつ．

本論文では，まず，**2.** において $tf \cdot idf$ を説明し，**3.** では本手法と $tf \cdot idf$ の違いと計算方法を述べる．**4.**，**5.**，**6.** では本手法に基づいて構築した対話処理による評価実験の結果について述べる．最後に，本手法の有効性と今後の課題を述べる．

2. $tf \cdot idf$

情報検索において，索引語の重み付けをするために $tf \cdot idf$ が用いられている．$tf \cdot idf$ について表 1 を用いて説明する．各行は文書集合中の文書，各列は索引語に対応する．d_1 の行は文書 d_1 に含まれる各単語の出現頻度である．t_1 の列は各文書に含まれる t_1 の単語数である．idf を式 (1) に示す．

$$idf(t) = \log \frac{N}{df(t)} + 1 \tag{1}$$

ここで，1 を加える理由は $\log \frac{N}{df(t)}$ が “0” の場合に対処するためである．求められた tf と idf との乗算が各単語の重みとなる．各文書における $tf \cdot idf$ の結果を表 2 に示す．情報検索を想定して，入力された単語に対して各文書における $tf \cdot idf$ の合計が大きい文書順に提示する処理を考える．検索を行う単語が，ターム t_1 とターム t_2 の場合，各文書における $tf_1 \cdot idf_1 + tf_2 \cdot idf_2$ の結果を以下に示す．

d_1 ：$2.58 + 2.00 = 4.58$
d_2 ：$0.00 + 3.00 = 3.00$
d_3 ：$5.16 + 1.00 = 6.16$

表 1　tf, df, idf の例
Table 1　Example of tf, df, idf.

	t_1	t_2	t_3	t_4
d_1	1	2	0	2
d_2	0	3	2	0
d_3	2	1	1	0
df	2	3	2	1
	$\log_2 \frac{D}{df} + 1$			
idf	2.58	1.00	1.58	2.58

表 2　$tf \cdot idf$ の例
Table 2　Example of $tf \cdot idf$.

	$tf_1 \cdot idf_1$	$tf_2 \cdot idf_2$	$tf_3 \cdot idf_3$	$tf_4 \cdot idf_4$
d_1	1×2.58	2×1.00	0×1.58	2×2.58
d_2	0×2.58	3×1.00	2×1.58	0×2.58
d_3	2×2.58	1×1.00	1×1.58	0×2.58

この結果から，t_1 と t_2 の検索語に対して，d_3，d_1，d_2 の順で選ばれる．次章では，$tf \cdot idf$ との違いを含め，情報量を用いた重み付けによる類似度の説明をする．

3.　$tf \cdot AoI$ の重み付けに基づく類似性

3.1　基本的な考え方

本手法は 2 文間の一致率による比較を拡張した手法であり，少ないデータ数においても類似度の計算が行える．一致率により比較する場合，各単語の重みに違いはないが，本手法は各単語に対して重み付けを行う点が違いである．本手法では音声対話処理を考慮して，間投詞などの冗長な表現が含まれるデータを処理する．これらの表現は個人に依存することが多いため，あらかじめ削除リストを作成するのは困難である．本手法では，実際のデータに基づいて単語の重み付けを行うことにより，この問題を解決する．更に，音声対話には，頻繁に倒置表現が含まれる．この問題には，単語の並びを無視した $tf \cdot idf$ の考え方を取り入れることで頑健に処理する．

ところで，音声対話処理は多くの基礎研究を必要とする．しかしながら，複数の研究を含めると，本手法の有効性及び問題点を明らかにできない場合がある．このような理由から，本論文ではユーザの発話に焦点を当て，話し言葉における質問文の同定を扱う問題とした．まず，質問文と応答文のペアがあらかじめ用意されている場合を想定する．質問の表現は数多く存在するため，入力文がコーパス中の質問文と完全一致することは少ない．そこで，コーパスから質問文と最も類似している文の選択を行い，正誤を判定する．

3.2　計算方法

本手法では，ベクトル空間モデルの表現形式を利用してユークリッド距離を計算する．2 次元のデータ (α, β)，(γ, δ) を考えた場合，ユークリッド距離は $\sqrt{(\alpha-\gamma)^2 + (\beta-\delta)^2}$ となる．我々はユークリッド距離を測定する要素として，「文における単語頻度 × 情報量」を使う．これを，$tf \cdot AoI$（term frequency × Amount of Information）と呼ぶ．

表 3 を用いて $tf \cdot AoI$ 及び，ユークリッド距離の算出方法と対話処理への適用について説明する．まず，基準となる文（以降，基準文とする）に対して単語の分割を行う．単語分割は茶筅 [11] を利用しているが，本手法では品詞情報は利用していない．これは，音声対話において単語分割をする場合，韻律情報を利用する方法 [14], [15] も考えられるため，他言語への適用を考慮して特定の言語に依存した知識を最小限としたためである．単語分割された基準文の「異なり単語数」を n として，n 次元のデータとする．分割された単語がその文中に出現する頻度（term frequency 以降 tf）を求める．基準文中に出現する「異なり単語」を $(t_1, t_2, \cdots, t_n)$ とした場合，tf は $(tf_1, tf_2, \cdots, tf_n)$ になる．基準文中の出現単語数だけを要素とするため，基準文の tf はすべて出現頻度が 1 回以上となる．重み付けした値，すなわちは $tf \cdot AoI$ は，$(tf_1 \cdot AoI(t_1), tf_2 \cdot AoI(t_2), \cdots, tf_n \cdot AoI(t_n))$ になる．ここで，類似度を測定する対象となる文（以降，対象文とする）を $Sent_A$ とする．対象文 $Sent_A$ の tf_a は $(tf_{a1}, tf_{a2}, \cdots, tf_{an})$ である．基準文と対象文の重み付けされた $tf \cdot AoI$ の値からユークリッド距離を計算する．ユークリッド距離を式 (2) に示す．

$$D = \sqrt{\sum_{i=1}^{n} (tf_i \cdot AoI(t_i) - tf_{ai} \cdot AoI(t_{ai}))^2} \quad (2)$$

表 3 の N はデータ中に存在する総単語数である．類似度を式 (3) に示す．

$$Similarity = \frac{1}{D+1} \quad (3)$$

2 文が同一文である場合，ユークリッド距離の値は 0 になるため，最も高い類似度は 1 となる．

次に表 4 の例文を用いて類似度の計算を行う．ここで，基準文は「バドミントンの時間です」として，比較する対象文は 3 文（$Sent_A$，$Sent_B$，$Sent_C$）とする．対象文を以下に示す．

表 3　類似度の求め方
Table 3　How to calculate a similarity.

基準文	$(t_1$	$, t_2$	$, \cdots\cdots ,$	$t_n)$
tf	$(tf_1$	$, tf_2$	$, \cdots\cdots ,$	$tf_n)$
AoI	$(-\log_2 \frac{tf_1}{N}$	$, -\log_2 \frac{tf_2}{N}$	$, \cdots\cdots ,$	$-\log_2 \frac{tf_n}{N})$
$tf \cdot AoI$	$(tf_1 \cdot AoI(t_1)$	$, tf_2 \cdot AoI(t_2)$	$, \cdots\cdots ,$	$tf_n \cdot AoI(t_n))$
対象文 $Sent_A$	$(t_{a1}$	$, t_{a2}$	$, \cdots\cdots ,$	$t_{an})$
tf_a	$(tf_a 1$	$, tf_a 2$	$, \cdots\cdots ,$	$tf_a n)$
$AoI(t_a)$	$(-\log_2 \frac{tf_a 1}{N}$	$, -\log_2 \frac{tf_a 2}{N}$	$, \cdots\cdots ,$	$-\log_2 \frac{tf_a n}{N})$
$tf_a \cdot AoI(t_a)$	$(tf_{a1} \cdot AoI(t_{a1})$	$, tf_{a2} \cdot AoI(t_{a2})$	$, \cdots\cdots ,$	$tf_n \cdot AoI(t_{an}))$
ユークリッド距離 D	$\sqrt{\sum_{i=1}^{n} (tf_i \cdot AoI(t_i) - tf_{ai} \cdot AoI(t_{ai}))^2}$			
類似度 $Similarity$	$\frac{1}{\sqrt{\sum_{i=1}^{n} (tf_i \cdot AoI(t_i) - tf_{ai} \cdot AoI(t_{ai}))^2 + 1}}$			

表 4　類似度の計算例
Table 4　Examples of similarity calculation.

基準文	バドミントンの時間です			
基準文の単語	(バドミントン	, の	, 時間	, です)
tf	(1	, 1	, 1	, 1)
対話例中の出現頻度	(1	, 164	, 14	, 11)
AoI	$(-\log_2 \frac{1}{4317}$	$, -\log_2 \frac{164}{4317}$	$, -\log_2 \frac{14}{4317}$	$, -\log_2 \frac{11}{4317})$
$tf \cdot AoI$	$(1 \times -\log_2 \frac{1}{4317}$	$, 1 \times -\log_2 \frac{164}{4317}$	$,1 \times -\log_2 \frac{14}{4317}$	$,1 \times -\log_2 \frac{11}{4317})$
対象文 $Sent_A$	病院の時間です			
基準文との一致数 (tf_a)	(0	, 1	, 1	, 1)
$tf_a \cdot AoI$ の式	$(0 \times -\log_2 \frac{1}{4317}$	$, 1 \times -\log_2 \frac{164}{4317}$	$,1 \times -\log_2 \frac{14}{4317}$	$,1\times -\log_2 \frac{11}{4317})$
$tf_a \cdot AoI$ の結果	(0	,0.30	,8.26	,8.61)
ユークリッド距離 D	$\sqrt{(12.07-0)^2 + (0.30-0.30)^2 + (8.26-8.26)^2 + (8.61-8.61)^2} = 12.07$			
対象文 $Sent_A$ との類似度 $Similarity$	$\frac{1}{12.07+1} = 0.0765$			
基準文の単語数を分母とした一致率	$\frac{3}{4} = 0.75$			
対象文 $Sent_B$	バドミントンの時間だ			
基準文との一致数 (tf_b)	(1	, 1	, 1	, 0)
$tf_b \cdot AoI$ の式	$(1 \times -\log_2 \frac{1}{4317}$	$, 1 \times -\log_2 \frac{164}{4317}$	$,1\times -\log_2 \frac{14}{4317}$	$,0 \times -\log_2 \frac{11}{4317})$
$tf_b \cdot AoI$ の結果	(12.07	,0.30	,8.26	,0)
ユークリッド距離 D	$\sqrt{(12.07-12.07)^2 + (0.30-0.30)^2 + (8.26-8.26)^2 + (8.61-0)^2} = 8.61$			
対象文 $Sent_B$ との類似度 $Similarity$	$\frac{1}{8.61+1} = 0.1041$			
基準文の単語数を分母とした一致率	$\frac{3}{4} = 0.75$			
対象文 $Sent_C$	私のバドミントンのサークル			
基準文との一致数 (tf_c)	(1	, 2	, 0	, 0)
$tf_c \cdot AoI$ の式	$(1 \times -\log_2 \frac{1}{4317}$	$, 2 \times -\log_2 \frac{164}{4317}$	$,0 \times -\log_2 \frac{14}{4317}$	$,0 \times -\log_2 \frac{11}{4317})$
$tf_c \cdot AoI$ の結果	(12.07	,0.60	,0	,0)
ユークリッド距離 D	$\sqrt{(12.07-12.07)^2 + (0.30-0.60)^2 + (8.26-0)^2 + (8.61-0)^2} = 11.94$			
対象文 $Sent_C$ との類似度 $Similarity$	$\frac{1}{11.94+1} = 0.0773$			
基準文の単語数を分母とした一致率	$\frac{2}{4} = 0.50$			

対象文 $Sent_A$ 「病院の時間です」
対象文 $Sent_B$ 「バドミントンの時間だ」
対象文 $Sent_C$ 「私のバドミントンのサークル」

この例で利用した出現頻度は，大学生の女性 3 名（話者 Sp_1，話者 Sp_2，話者 Sp_3）の対話例から求めた単語の出現頻度である．対話データは，2 名の自由発話を書き起こしたデータである．女子大学生 3 名から 2 名ずつ（$Sp_1 \leftrightarrow Sp_2$，$Sp_1 \leftrightarrow Sp_3$，$Sp_2 \leftrightarrow Sp_3$）に対話をしてもらい，書き起こした．この書き起こしデータは，一般に公開しているデータではなく，3 名の大学生による 3 時間程度の会話を収集したデータである．この例は，書き起こしたデータに含まれる単語から生成した文である．基準文を構成している単語の出現回数は「バドミントン (1) の (164) 時間 (14) です (11)」である．ここで括弧内は前単語の対話データ中の出現回数を示す．本手法では，基準文以外で現れた単語は考慮しないため，対象文に出現する「病院 (1), だ (85), 私 (12), サークル (2)」は扱わない．基準文の tf は基準文に出現する頻度であり，各単語が 1 回ずつ出現しているので (バドミントン, の, 時間, です)=(1,1,1,1) と

なる．対象文 $Sent_A$ の tf は基準文の単語が出現した回数であり，(バドミントン, の, 時間, です)=(0,1,1,1) となる．対象文 $Sent_C$ の tf では「の」が 2 回出現しているので，(バドミントン, の, 時間, です)=(1,2,0,0) となる．ここでは出現順序を考慮せずに，出現頻度のみを扱う．計算に利用される対話データの総単語数 N は 4,317 である．$AoI(t_i)$ は $-\log_2 \frac{tf_i}{N}$ で計算するので，基準文の AoI は ($-\log_2 \frac{1}{4317}$, $-\log_2 \frac{164}{4317}$, $-\log_2 \frac{14}{4317}$, $-\log_2 \frac{11}{4317}$) となる．この AoI に基づいて，各対象の $tf \cdot AoI$ を求める．対象文 $Sent_A$ の $tf \cdot AoI$ は (0, 0.30, 8.26, 8.61)，対象文 $Sent_B$ の $tf \cdot AoI$ は (12.07, 0.30, 8.26, 0)，対象文 $Sent_C$ の $tf \cdot AoI$ は (12.07, 0.60, 0, 0) である．基準文と各対象文との類似度を求めると，対象文 $Sent_A$ は 0.0765，対象文 $Sent_B$ は 0.1041，対象文 $Sent_C$ は 0.0773 である．したがって，$Sent_B > Sent_C > Sent_A$ となり対象文 $Sent_B$ との類似度が最も高くなり，次いで $Sent_C$，$Sent_A$ の順に高いという計算結果となる．この関係は類似性を正しくとらえていると考えられる．基準文の単語数を分母に一致率を求めると，対象文 $Sent_A$ は 0.75，対象文 $Sent_B$ は 0.75，対象文 $Sent_C$ は 0.50 となる．関係は対象文 $Sent_A$ と対象文 $Sent_B$ の一致率が等しく，次いで $Sent_C$ となる．上記の結果から，$tf \cdot AoI$ の重み付けによる類似度は一致率より有効と考えられるため，本論文では比較実験を行い，有効性を確認する．

4. 予備実験

4.1 目　　的

ATR の SLDB [16] のデータに対して，交差検定法による評価を行い，他手法との比較を行う．ここでは，下記の 3 手法で比較実験を行う．

- $tf \cdot AoI$ の重み付けによる類似度
- $tf \cdot idf$ の重み付けによる類似度
- 入力文の単語数を分母とした一致率

“$tf \cdot AoI$ の重み付けによる類似度” は本手法である．本実験では，入力文を基準文として計算を行う．“$tf \cdot idf$ の重み付けによる類似度” は単語の重み付け以外は本手法と同様である．“入力文の単語数を分母とした一致率” は一致率の最も高い質問文に対する応答を出力する．

4.2 実験方法

ATR の SLDB [16] のデータ 12,095 文から，ホテルの受付の発話文とお客の発話文の対（つい）を収集した．その結果，5,812 対が選択され，総単語数は 266,331 語，異なり単語は 5,078 単語のデータとなった．実験データの作成において，ホテルの受付の発話文とお客の発話文の数が一致しない場合，最後に発話された発話文を削除した．その結果，削除された文は $12,095 - (5,812 \times 2) = 471$ 文である．実験は，交差検定法により行った．すなわち，5,812 対から 1 対を入力文と入力文に対する正解応答とみなし，残りの 5,811 対を訓練データとして，これを 5,812 回繰り返す．そのため，オープンデータの実験となる．訓練データから入力文と最も類似している文を見つけ，その対である応答を選択する．入力文に対する正解応答と同一表現の場合，正応答とする．

表 5　正応答数
Table 5　Correct response number of the same expression.

	$tf \cdot AoI$(本手法)	$tf \cdot idf$ の重み	一致率
正応答数	152	147	123

4.3 実験結果と考察

表 5 は 5,812 文の入力に対する正応答数である．実験結果から，本手法の正応答数は 152 応答となり，最も高い数値となった．5,812 応答において，正解が 152 応答であり，低い結果となった．この原因は，5,811 対の訓練データ中に正応答が存在しないためである．訓練データ中に存在しない応答数が，5,812 文中 5,084 文存在した．つまり，訓練データに正解が存在したのは，$5,812 - 5,084 = 728$ 文であった．728 応答中に正解応答とは表現が異なるが意味が近い応答が存在したが，正応答の基準があいまいとなるため，**5.** において，実験データの選択を考慮した実験を行う．

5. 評価実験

5.1 目　　的

実験の目的は，本手法と他手法の比較実験を行い，本手法の有効性を明らかにすることである．表 3 の計算方法により類似度計算を行う本手法と「$tf \cdot idf$ の重み付けによるユークリッド距離による類似度」及び，「入力文の単語数を分母とした一致率」による手法との比較を行う．

5.1.1 実験データの収集

本手法は音声対話処理を最終対象としている．そのため，音声対話を書き起こした ATR の SLDB [16] を利用する．公平さを保つために，既存のコーパスから

データを収集する．ここでは SLDB のデータ 12,095 文から，意味が等しく異なる表現の 2 文とそれに対応した応答を収集する．図 1 を用いて獲得手順を説明する．

（1） 図 1 の A2 と AN のように，表現が完全に一致する文を探す

（2） 前文の表現が異なる文であることを確かめる

（3） 前文が異なる表現である場合，2 文以上 4 文以下の組をデータ候補として抽出

（4） 候補データから意味が等しい文であれば訓練データと評価データとして保存

ここで，意味が等しい文とは，男子大学生 4 名及び大学院生 4 名の合計 8 名により半数以上が等しいと評価した文である．上記の条件で獲得されたデータは 12,095 文から 154 セットであった．異なり単語数は 707，単語数は 4,687 のデータである．表 6 に収集データの例を示す．質問 1 と応答が訓練データであり，質問 2 がテストデータとなる．データの都合上，ここで質問文と呼んでいる文が必ずしも疑問文になるとは限らない．ここで 1 セットは，表現が異なり意味が等しい 2 文とそれに対する同一の応答である．収集データは，訓練データとテストデータに同じ表現は存在しないため，完全なオープンデータでの実験となる．

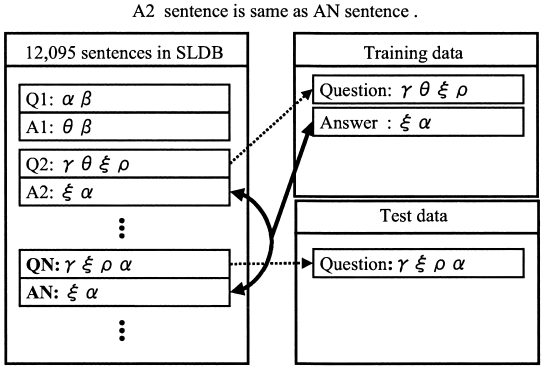

図 1 データ収集方法
Fig. 1 How to collect data.

5.2 実験方法

図 2 を用いて実験方法を説明する．本手法による類似度の測定を対話処理に応用する．あらかじめ質問応答のセットを用意し，入力文と最も類似している質問文に対する応答を出力とする．訓練データは 173 組からなり，前節で収集したデータである．収集された

表 6 収集データの例
Table 6 Example of an acquired data.

質問 1	ええ、結構でございますが、あただ、チェックアウトは午前十時となっております。 それ以降の御利用といいますのは、追加料金一時間で二千円をいただいております。 でですから、二時までですと四時間ということになりますので、 ちょうど合計八千円となりますが、よろしいでしょうか。
質問 2	ええ、結構でございます。ただしチェックアウトは午前十時となっております。 そしてそれ以後の料金の方なんですけれども、御利用いただきますのに一時間二千円ということで 追加料金をいただいております。ですから午後二時までですと、四時間となりますので、 えー合計八千円ということになりますが、よろしいですか。
応答	ええ、結構です。じゃあ延長ということでお願いします。
質問 1	近代美術館のほうはロックフェラーセンターの一ブロック北側にあります。 ロックフェラーセンターの場所はご存じですか。
質問 2	近代美術館はロックフェラーセンターの一ブロック北側にございます。 ロックフェラーセンターの場所はご存じですか。
応答	はい、分かります。
質問 1	先ほどチェックインしました百七号室の鈴木和子です。
質問 2	先ほどチェックインしました一〇七号室の鈴木和子と申しますが。
応答	はい、鈴木様、どういった御用件でしょうか。
質問 1	はい、承知いたしました。御宿泊のお客様のお名前をどうぞ。
質問 2	かしこまりました、早速確認いたします。 お泊まりのお客様のお名前を、おっしゃっていただけますか。
応答	はい、山下正博さんという方です。
質問 1	そうですか、分かりました。いろいろと教えてくれてありがとう。
質問 2	そうですか、いろいろと教えていただいてありがとう。お手数をおかけします。
応答	ありがとうございました。
質問 1	朝食は別になります。一人当たり十一ドルになります。ホテル内のレストランではどこでも結構です。
質問 2	えご予約はできますけれども、料金は別となっております。ホテル内のレストランはどこでもご利用になれます。 そして料金は十一ドルとなっております。
応答	分かりました。じゃあよろしくお願いします。

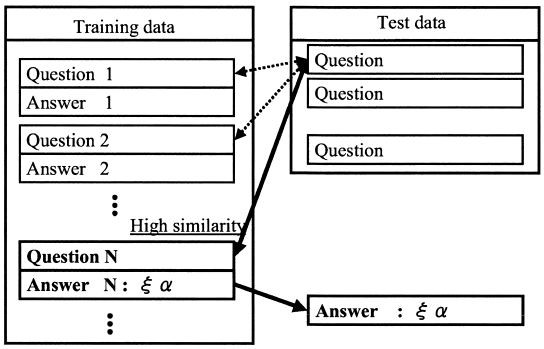

図 2 実験方法
Fig. 2 Experiment procedure.

表 7 実験結果
Table 7 Experiment results.

	本手法	$tf \cdot idf$ の重み	一致率
正解応答数	107/154	102/154	92/154
正解応答率	69.5%	66.2%	59.7%
本手法では誤応答となる正応答数	-	2	1

データは，異なる表現の質問が二つ存在するため，訓練データに存在しないもう一つの質問をテストデータとした．図 2 のように，テストデータの質問文と訓練データ中の各質問文の類似性を計算し，最も類似性の高い質問に対応している応答を出力する．本実験では，**4.2** の実験方法と同様の 3 手法による比較実験を行う．

5.3 実験結果

実験結果を表 7 に示す．本手法は 154 の質問のうち 107 個の正しい選択を行った．本手法の正応答率では 69.5%，一致率では 59.7%となり，9.8 ポイント上回った．“$tf \cdot idf$ の重み付けによる類似度” では 66.2%となり，本手法が 3.3 ポイント上回った．本手法では誤応答となるが，“入力文の単語数を分母とした一致率” による選択では正解となる応答数は 1 応答であった．本手法では誤応答となるが，“$tf \cdot idf$ の重み付けによる類似度” による選択では正解となる応答数は 2 応答であった．本手法と他手法において正応答の選択における差は少ないことから，本手法は他の手法が正応答を選択する性質をほぼ包含している．

5.4 考　　察

表 8 の実験結果の例を用いて，本手法で選択した場合と “一致率” で選択した場合との違いについて述べる．入力文は「はい、三人部屋でしたらございますが、あしかしえーお子様の年齢はおいくつですか。」とする．この文が基準文となり，文中には音声対話特有の表現が含まれ,「**あ**しかし**えー**」の太字単語は間投詞である．本手法が最も類似していると選択した文（正解），“えお子様はお幾つでいらっしゃいますか” を対象文 $Sent_A$ とする．入力文の単語数を分母とした一致率により選択した文,「ええ、結構でございますが、あただ、チェックアウトは午前十時となっております。それ以降の御利用といいますのは、追加料金一時間で二千円をいただいております。でですから、二時までですと四時間ということになりますので、ちょうど合計八千円となりますが、よろしいでしょうか。」を対象文 $Sent_B$ とする．一致率は入力文の単語数を分母として，対象文との一致数を求める．

$$\text{一致率} = \frac{\text{入力文の各単語数との一致数}}{\text{入力文の単語数}} \qquad (4)$$

ただ，対象文のある単語 A が入力文の単語 A の個数を超えていても，最大値は入力文の個数とする．例えば，表 8 における対象文 $Sent_B$ の入力文の単語と一致している単語（下線単語）は 26 単語あるが，11 単語が一致した単語数となる．対象文 $Sent_A$ に対して一致率は $\frac{6}{23} = 26.1\%$ となり，対象文 $Sent_B$ に対しては $\frac{11}{23} = 47.82\%$ であった．ここで，本手法における対象文 $Sent_A$ との類似度の結果を説明する．属性「はい」の入力文の頻度は “1” であり，対象文 $Sent_A$ の頻度は “0” である．総単語数が 4,687 回で,「はい」の出現回数が 75 回なので,「はい」の情報量は $-\log_2 \frac{75}{4687}$ になる．計算式は以下のようになる．
$\sqrt{((1-0)\cdot -\log_2 \frac{75}{4687})^2 + ((2-0)\cdot -\log_2 \frac{237}{4687})^2 +}$
対象文 $Sent_B$ に対しての計算式は以下のようになる．
$\sqrt{((1-0)\cdot -\log_2 \frac{75}{4687})^2 + ((2-7)\cdot -\log_2 \frac{237}{4687})^2 +}$
本手法では対象文 $Sent_A$ との類似度が 0.0270(36.09)(注1)，対象文 $Sent_B$ との類似度が 0.0210(46.57) となり，単語の重みを考慮することで正しい選択が行われた．一致率の場合，最も高い一致率が複数存在することが多く，更なる選択手法が必要となる．しかしながら，本手法では各文に対して計算結果が等しくなることがなく，選択が容易である．

本手法における誤応答は，154 文中 46 文であった．本実験の評価では，類似度の順位が 1 位以外を誤りとして扱った．そのため，類似度の順位が 2 位であったために誤りとなった例は，10 文存在した．更に，類似度の順位 2〜5 位に正解が含まれれる場合は 25 文であり，誤りの 54.34%(=25/46) を占めていた．ここで，

(注1)：括弧内はユークリッド距離.

表 8　応答文の選択が異なる例
Table 8　Different example of a response selection.

入力文	はい、三人部屋でしたらございますが、あしかしえーお子様の年齢はおいくつですか。																					
対象文 $Sent_A$	え お子様はお 幾つでいらっしゃい ます か。																					
対象文 $Sent_B$	ええ 、結構 で ございます が 、あ ただ 、チェック アウト は 午前 十 時 と なっ て おり ます 。それ 以降 の 御 利用 といいます の は 、追加 料金 一 時間 で 二 千 円 を いただい て おり ます 。で です から 、二 時 まで です と 四 時間 という こと に なり ます ので 、ちょうど 合計 八 千 円 と なり ます が 、よろしい でしょう か 。																					
属性	はい	、	三	人	部屋	でし	たら	ござい	ます	が	あ	しかし	えー	お子様	の	年齢	は	お	いくつ	です	か	。
$f(x)$	75	237	40	12	1	4	9	90	199	66	6	0	11	1	131	1	76	56	0	140	110	476
入力文の頻度	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$Sent_A$ との一致数	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	1
$Sent_B$ との一致数	0	7	0	0	0	0	0	1	5	2	1	0	0	0	2	0	2	0	0	2	1	3
$Sent_A$ 本手法	$\frac{1}{\sqrt{((1-0)\times-\log_2\frac{75}{4687})^2+((2-0)\times-\log_2\frac{237}{4687})^2+\cdots+1}} = \frac{1}{36.09+1} = 0.0270$																					
$Sent_A$ 一致率	$\frac{6}{23} = 26.1\%$																					
$Sent_B$ 本手法	$\frac{1}{\sqrt{((1-1)\times-\log_2\frac{75}{4687})^2+((2-7)\times-\log_2\frac{237}{4687})^2+\cdots+1}} = \frac{1}{46.57+1} = 0.0210$																					
$Sent_B$ 一致率	$\frac{11}{23} = 47.82\%$																					

2 位以内に含まれた文の例を挙げる．質問文が「四名様でご予約ですね。」であった場合，本手法では類似度が最も高い文として,「はい、ジョン・フィリップス様ですね。何名様でご予約でしょうか。」を選択した．質問文に対する正解とされる質問文はランキング 2 位として計算されており,「四名様でございますね。」である．このような問題に対しては，質問文の単語を基準としているため，対象文の単語に含まれる大きな重みの単語が不一致であることを考慮していないことが原因と考えられる．そこで，**6.** では対象文の単語を考慮した実験を行う．更に，類似度の順位が 5 以内に含まれない誤りのパターンは，45.66%(=21/46) であり，大きな重み付けされている単語が一致していない文であった．この誤りに対しては，同義語の言換えが必要と考えられる．この件に関しては本手法では扱わないが，今後扱うことを考えている．

6. 提案手法のパラメータ評価実験

6.1 目　　的

$tf \cdot AoI$ の重み付けによる類似度では，基準文の設定や係数の付与により計算結果が異なる．本実験では，基準文の設定と係数の値を変更することで精度向上を試みる．まず，基準文の設定によるユークリッド距離の違いは下記のように D_1，D_2 の二つが考えられる．

D_1 ：入力文を基準文とした場合

D_2 ：コーパス中の各質問文を基準文とした場合

前章の評価実験では D_1 による計算であり，入力文を基準文としてコーパスに含まれる各質問文を対象文としている．それに対して，D_2 ではコーパスに含まれる各質問文を基準文とし，入力文を対象文として類似度を計算する．例えば，例文 1「僕の学校」と例文 2「僕の学校だ」の類似度を計算するには．例文 1 を基準文として例文 2 を比較した場合，属性数は (僕, の, 学校) の三つとなり，例文 2 を基準文として例文 1 を比較した場合，属性数は (僕, の, 学校, だ) の四つとなる．つまり，基準文を変えることで，属性の設定が異なり，類似度の計算結果にも影響する．このように類似度の計算結果も異なることから，D_1 と D_2 の計算結果の合計に係数 α を付与した式を $D_1 + \alpha \times D_2$ として，この係数 α の最適値を求める．

表 9　実 験 結 果
Table 9　Experiment results.

	D_1	D_2	$D_1 + D_2$
正解応答数	107/154	30/154	86/154
正解応答率	69.5%	19.5%	55.8%
本手法では誤応答となる正応答数	-	7	9

6.2 実験結果と考察

まず，基準文の設定による実験結果を表 9 に示す．参考として $D_1 + \alpha \times D_2$ の $\alpha = 1$ の場合も付け加えた．D_1 の正応答率が，最も良い結果を示した．D_2 が 154 応答中 30 の正応答であり，最も低い結果となった．しかしながら，正応答の 30 応答中に「D_1 では

表 10　計 算 方 法
Table 10　How to calculate.

入力文	はい、十三日の六時ぐらいですね。何名様でお越しでしょうか。	D_1	D_2	D_1 + D_2
質問 1	はい、八月十三日の六時ぐらいでございますね。何名様でしょうか。	0.0807	0.0747	0.0406
応答 1	大人 四 人 です 。	類似度高い	類似度低い	類似度高い
質問 2	はい 、何 名 様 でしょ う か 。	0.0388	0.2326	0.0344
応答 2	三 人 です 。	類似度低い	類似度高い	類似度低い
入力文=基準文	(はい,、, 十, 三, 日, の, 六, 時, ぐらい, です, ね,。, 何, 名, 様, で, お越し, でしょ, う, か)			
入力文	(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1)			
質問 1	(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 2, 1, 1, 1, 1, 0, 1, 1, 1)			
質問 2	(1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1)			
質問 1=基準文	(はい,、, 八月, 十, 三, 日, の, 六, 時, ぐらい, で, ござい, ます, ね,。, 何, 名, 様, でしょ, う, か)			
質問 1	(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1)			
入力文	(1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 2, 1, 1, 1, 0, 1, 1)			
質問 2=基準文	(はい,、, 何, 名, 様, でしょ, う, か,。)			
質問 2	(1, 1, 1, 1, 1, 1, 1, 1, 1)			
入力文	(1, 1, 1, 1, 1, 1, 1, 1, 2)			

誤応答であり，D_2 では正応答」になる質問が七つ存在した．正応答が 30 応答と少ないにもかかわらず，7 応答が “D_1” とは異なる正応答であった．これは，D_1 とは選択傾向が他手法と比較しても異なることが分かる．この結果から，$D_1 + \alpha \times D_2$ の係数 α を変更することにより向上が可能と考えられる．

表 10 を用いて D_1，D_2，$D_1 + D_2$ の計算方法と結果の違いを説明する．入力文が “はい、十三日の六時ぐらいですね。何名様でお越しでしょうか” の場合を考える．入力文を基準とした場合，属性は (はい,、, 十, 三, 日, の, 六, 時, ぐらい, です, ね,。, 何, 名, 様, で, お越し, でしょ, う, か) であり，入力文のベクトルは (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1) になる．質問文 1 は “はい、八月十三日の六時ぐらいでございますね。何名様でしょうか。” なので，入力文を基準とした場合，=(1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 2, 1, 1, 1, 1, 0, 1, 1, 1) になる．質問文 1 を基準文とした場合，属性は (はい,、, 八月, 十, 三, 日, の, 六, 時, ぐらい, で, ござい, ます, ね,。, 何, 名, 様, でしょ, う, か) であり，質問文のベクトルは (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1) になる．入力文のベクトルは (1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 2, 1, 1, 1, 0, 1, 1) である．D_1 の場合，質問文 1 との類似度は 0.0807，質問文 2 は 0.0388 であり，質問 1 の類似度が高い．D_2 の場合，質問文 1 との類似度は 0.0747，質問文 2 は 0.2326 であり，質問文 2 の類似度が高い．$D_1 + D_2$ の場合，質問文 1 は 0.0406，質問文 2 は 0.0344 であり，質問文 1 の類似度が高い．この例では，D_1 と $D_1 + D_2$ の選択結果は等しいが，

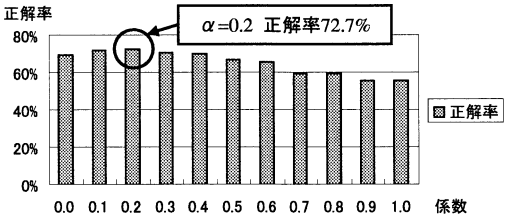

図 3　α の評価実験
Fig. 3　α Evaluation experiment.

$D_1 + D_2$ は D_2 の結果が影響し，正応答率が低下したと考えられる．質問 2 のような短い文を基準文とする場合，要素数が減少し，ノイズに対する耐性が低下する傾向にある．D_2 では，要素数が少なくなるため，質問 2 が質問 1 よりも類似度が高くなり，誤応答となった．$D_1 + D_2$ による選択方法は，D_2 の値を加えることで，比較対象となる文の要素数を考慮することになる．

次に，$D_1 + \alpha \times D_2$ の係数 α の最適値を求める．正応答率を向上させるために，$D_1 + \alpha \times D_2$ の式において，α の値を 1 より小さい値で変動させる評価実験を行った．その実験結果を図 3 に示す．$\alpha = 0.2$ のときに，正応答率が 72.7%となり最も高い値を示した．表 7 の結果と比較した場合，D_1 に対して 3.3 ポイント，“$tf \cdot idf$ の重み付けを用いたユークリッド距離に基づく類似度” に対して 6.5 ポイント，“入力文の単語数を分母とした一致率” に対して 13 ポイントの向上が確認された．$D_1 + \alpha \times D_2$ の係数 α を変更することが有効であることが確認された．

今回，$\alpha = 0.2$ に最適な係数が存在したが，いかな

る場合でも $\alpha = 0.2$ が最適とはいえない．他のデータの場合でも，最適な値であるかを確認する必要がある．しかしながら，日本語における話し言葉の対話データは少なく，本論文における実験データのような異なる表現で同等の意味をもつ質問文を含んでいるデータの収集が困難であるため，本実験で利用した 154 セットを二つに分割した．ここで，分割されたデータにおいても係数 α の最適値が一定であるのか，確認する．図 4 の前半 77 セットに対する正解率，図 5 の後半 77 セットに対する正解率において，係数 $\alpha = 0.2$ が最も高い正解率となった．この結果から，データにおける最適値は変化しないことが確認された．これは，少量のデータからでも最適な係数を予測可能であるといえる．

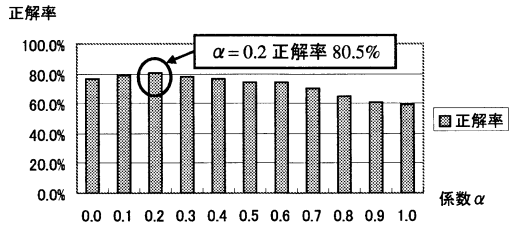

図 4　154 文の前半 77 文の係数 α の評価実験
Fig. 4　Evaluation experiment in the first half of experiment data.

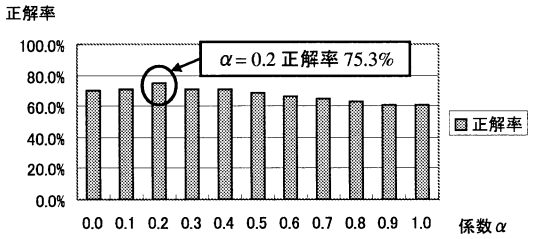

図 5　154 文の後半 77 文の係数 α の評価実験
Fig. 5　Evaluation experiment in the latter half of experiment data.

係数の変化による類似度の変化を説明する．表 11 に評価データとして入力した文,「はい、三人部屋でしたらございますが、あしかしえーお子様の年齢はおいくつですか。」の計算結果を示す．ここでは，各しきい値 α における，類似度の高い文を降順に表示する．

ここで，係数 α の特徴について説明する．表 11 において，対象文を考慮しない場合（$\alpha = 0.0$）では，対象文が非常に長い文でも，一致する単語が多く含まれる文を類似性が高い文として評価した．例えば，表 11 の $\alpha = 0.0$ の 2 番目に類似度が高い文は,「かしこまりました。ビサインターナショナルで、四五四零零五三八八六一九四一二五でございますね。お車の方はいつご用意すればよろしいですか。」のように長い文となった．対象文を考慮した場合（$\alpha \neq 0.0$），一致する単語と一致しない単語を多く含む非常に長い文が選択されなくなった．そのため，表 11 の $\alpha = 0.1 \sim 0.4$ では，類似度が高いとされた上位の文は比較的短い文であり，正解文が最も類似度が高い文となった．しかしながら，係数 α が大きくなるに従って，出現頻度が高い単語から構成された短い文が高い類似度になった．これは，単語が一致していない場合，重みが小さい単語（出現頻度が高い単語）から構成されている文の類似性が高くなるためである．そのため，$\alpha \geq 0.5$ では不正解となった．この実験結果から，$\alpha \geq 0.5$ では，

表 11　係数の評価実験結果の例
Table 11　Example of the coefficient evaluation experiment.

質問文：はい、三人部屋でしたらございますが、あしかしえーお子様の年齢はおいくつですか。						
α	評価データの正解質問文	D_1	D_2	$D_1 + \alpha \cdot D_2$	類似度	正誤
0.0	えお子様はお幾つでいらっしゃいますか。	36.09	18.24	36.09	0.0270	正
0.0	かしこまりました。ビサインターナショナルで、四五四零零五三八八六一九四一二五でございますね。お車の方はいつご用意すればよろしいですか。	36.15	58.02	36.15	0.0269	誤
0.1	えお子様はお幾つでいらっしゃいますか。	36.09	18.24	37.914	0.0257	正
0.1	はい、えー何名様でいらっしゃいますか。	36.87	17.91	38.661	0.0252	誤
0.2	えお子様はお幾つでいらっしゃいますか。	36.09	18.24	39.738	0.0245	正
0.2	はい、えー何名様でいらっしゃいますか。	36.87	17.91	40.452	0.0241	誤
0.3	えお子様はお幾つでいらっしゃいますか。	36.09	18.24	41.562	0.0235	正
0.3	はい、えー何名様でいらっしゃいますか。	36.87	17.91	42.243	0.0231	誤
0.3	ありがとうございました。	39.3	10.35	42.405	0.0230	誤
0.4	えお子様はお幾つでいらっしゃいますか。	36.09	18.24	43.386	0.0225	正
0.4	ありがとうございました。	39.3	10.35	43.44	0.0225	誤
0.5	ありがとうございました。	39.3	10.35	44.475	0.0220	誤
0.5	えお子様はお幾つでいらっしゃいますか。	36.09	18.24	45.21	0.0216	正

表 12　Mean Reciprocal Rank の結果
Table 12　Results of Mean Reciprocal Rank.

手法	係数 α	RR	MRR
$Matching$	-	107.282	0.697
$tf \cdot idf$	-	113.765	0.739
$tf \cdot AoI$	0.0	117.600	0.764
$tf \cdot AoI$	0.1	122.422	0.795
$tf \cdot AoI$	0.2	123.475	0.802
$tf \cdot AoI$	0.3	120.730	0.784
$tf \cdot AoI$	0.4	118.898	0.772
$tf \cdot AoI$	0.5	114.566	0.744
$tf \cdot AoI$	0.6	111.539	0.724
$tf \cdot AoI$	0.7	102.776	0.667
$tf \cdot AoI$	0.8	102.776	0.667
$tf \cdot AoI$	0.9	98.565	0.64
$tf \cdot AoI$	1.0	97.178	0.631

$Matching$　入力文の単語数を分母とした一致率
$tf \cdot idf$　$tf \cdot idf$ の重み付けによる類似度
$tf \cdot AoI$　$tf \cdot AoI$ の重み付けによる類似度

出現頻度が高い単語から構成された短い文を過大に評価してしまう傾向にある．これらの結果から，データにより最適な係数は変化するが，$\alpha = 0.1 \sim 0.4$ に存在すると考えられる．

類似度の順位が計算可能な場合，他にも評価方法が存在する．Mean Reciprocal Rank（MRR）[1] はランキングによる評価であるため，類似度評価に適していると考えられるため，MRR による評価実験を行った．MRR は，検索質問で利用されている評価方法であり，最初に出現した正解の順位の逆数を求め，それらを全文にわたって平均する．MRR の式は下記のとおりである．

$$MRR = \frac{\sum_{i=1}^{n} RR_i}{n} \tag{5}$$

$$RR_i = \frac{1}{Rank_i} \tag{6}$$

n は質問セット数である．表 12 に MRR の結果を示す．$\alpha = 0.2$ において，MRR は最も高く，0.802 となり，MRR の評価においても，図 3 と同様の結果となった．

7. むすび

本論文では $tf \cdot AoI$ の重み付けによる文の類似性を測定する手法を提案した．更に，本手法を対話処理に適用する方法を説明し，評価実験を行った．**5.3** の比較実験では，本手法の正応答率では 69.5%，一致率では 59.7%となり，9.8 ポイント上回った．“$tf \cdot idf$ の重み付けを用いたユークリッド距離に基づく類似度” では 66.2%となり，本手法が 3.3 ポイント上回った．更に，本手法を向上させるために $D_1 + \alpha \times D_2$ による計算を行い，$\alpha = 0.2$ のときに，“入力文の単語数を分母とした一致率” に対して 13 ポイント，“$tf \cdot idf$ の重み付けを用いたユークリッド距離に基づく類似度” に対して 6.5 ポイント上回った．これらの結果から，本手法の有効性が確認された．

本論文では，同じ意味をもつ異なる表現の質問文の同定を行ったが，正解応答も複数存在するため，今後の評価方法として，正解とされる応答の異なる表現を緩和することを考えている．更に，本手法では文脈情報を利用していないため，文脈の類似度を含めることを予定している．他にも，対話の対象分野に適応させるために，局所的な情報をフィードバックを行い，実際の音声対話に適用させることを考えている．

文　献

[1] J. Fukumoto, T. Kato, and F. Masui, “Question answering callenge(QAC-1) an evaluation of question answering tasks at the NTCIR workshop 3,” Proc. AAAI Spring Symposium, pp.134–137, March 2003.

[2] 相澤彰子, “語と文書の共起に基づく特徴度の数量的表現について,” 情処学論, vol.41, no.12, pp.3332–3343, 2000.

[3] 徳永健伸, 言語と計算 5　情報検索と言語処理, 辻井潤一（編）, 東京大学出版会, 東京, 1999.

[4] J. Weizenbaum, “ELIZA—A computer program for the study of natural language comunication between man and machine,” Communications of the Association for Computing Machinery, vol.9, no.1, pp.36–45, 1966.

[5] Y. Takebayashi, H. Tsuboi, H. Kanazawa, Y. Sadamoto, H. Hashimoto, and H. Shinichi, “A real-time speech dialog system using spontaneous speech understanding,” IEICE Trans. Inf. & Syst., vol.E76-D, no.1, Jan. 1993.

[6] V. Zue, S. Seneff, J. Glass, J. Polifroni, C. Pao, T.J. Hazen, and L. Hetherington, “JUPITER: A telephone-based conversational interface for weather information,” IEEE Trans. Speech Audio Process., vol.8, no.1, Jan. 2000.

[7] D. Litman, S. Pan, and M. Walker, “Evaluating response strategies in a web-based spoken dialogue agent,” Proc. ACL/COLING 98, 1998.

[8] 伊藤克亘, 自然言語処理—基礎と応用, 田中穂積（監修）, pp.302–322, 電子情報通信学会, 東京, 1999.

[9] 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,” 自然言語処理, vol.10, no.4, pp.145–175, 2003.

[10] 辻　慶太, 芳鐘冬樹, 影浦　峡, “対訳コーパスにおける低頻度語の性質—訳語自動抽出に向けた基礎研究,” 情処学 NL 研報, NL-138, pp.47–54, 2000.

[11] 松本裕治, 北内　啓, 山下達雄, 平野善隆, 松田　寛, 高岡

一馬，浅原正幸，日本語形態素解析システム『茶筌』version 2.2.1 使用説明書，Dec. 2000.
[12] 相澤彰子，“低頻度語の利用によるテキスト分類性能の改善と評価，” 情処学論，vol.44, no.7, pp.1720–1730, 2003.
[13] 松井くにお，田中穂積，“初期質問文から蓄積された質問応答への効果的マッチング法，” 自然言語処理，vol.10, no.5, pp.121–138, 2003.
[14] 小松昭男，太平栄二，市川　熹，“韻律情報を利用した構文推定およびワードスポットによる会話音声理解方式，” 信学論（D），vol.J71-D, no.7, pp.1218–1228, July 1988.
[15] T. Ohsuga, Y. Horiuchi, and A. Ichikawa, “Estimating syntactic structure from prosody in Japanese speech,” IEICE Trans. Inf. & Syst., vol.E86-D, no.3, pp.558–564, March 2003.
[16] T. Morimoto, N. Uratani, T. Takezawa, O. Furuse, Y. Sobashima, H. Iida, A. Nakamura, Y. Sagisaka, N. Higuchi, and Y. Yamazaki, “A speech and language database for speech translation research,” Proc. ICSLP’94, pp.1791–1794, 1994.

（平成 16 年 1 月 9 日受付，3 月 5 日再受付）

木村　泰知　（学生員）

平 8 北海学園大・工・電子情報卒．平 13 同大大学院工学研究科修士課程了．現在，北大大学院工学研究科博士課程在学中．自然言語処理の研究に従事．

荒木　健治　（正員）

昭 57 北大・工・電子卒．昭 63 同大大学院博士課程了．工博．同年，北海学園大学工学部電子情報工学科助手．平元同講師．平 3 同助教授．平 10 同教授．平 10 北大・工・電子情報工学専攻助教授．平 14 同教授．現在，北大・情報科学・メディアネットワーク専攻教授．自然言語処理，特に機械翻訳，音声対話処理に関する研究に従事．情報処理学会，人工知能学会，言語処理学会，日本認知科学会，ACL，IEEE 各会員．

栃内　香次　（正員）

昭 37 北大・工・電気卒．昭 39 同大大学院工学研究科電気工学専攻修士課程了．現在，北海学園大学大学院経営学研究科教授．主として音声情報処理，自然言語処理の研究に従事．工博．情報処理学会，日本音響学会各会員．