

述語項構造を介した文の選択に基づく 音声対話用言語モデルの構築

Language Modeling for Spoken Dialogue System Based on Sentence Filtering Using Predicate-Argument Structures

吉野 幸一郎
Koichiro Yoshino

京都大学 情報学研究科
School of Informatics, Kyoto University
yoshino@ar.media.kyoto-u.ac.jp

森 信介
Shinsuke Mori

(同 上)
mori@ar.media.kyoto-u.ac.jp

河原 達也
Tatsuya Kawahara

(同 上)
kawahara@ar.media.kyoto-u.ac.jp

keywords: Spoken Dialogue System, Language Modeling, Predicate-Argument Structure

Summary

A novel text selection approach for training a language model (LM) with Web texts is proposed for automatic speech recognition (ASR) of spoken dialogue systems. Compared to the conventional approach based on perplexity criterion, the proposed approach introduces a semantic-level relevance measure with the back-end knowledge base used in the dialogue system. We focus on the predicate-argument (P-A) structure characteristic to the domain in order to filter semantically relevant sentences in the domain. Moreover, combination with the perplexity measure is investigated. Experimental evaluations in two different domains demonstrate the effectiveness and generality of the proposed approach. The combination method realizes significant improvement not only in ASR accuracy but also in semantic-level accuracy.

1. はじめに

これまで数多くの音声対話システムが研究開発され、一部は実世界で利用されるようになってきている。特に近年、スマートフォンなどで多様な要求に応答を行うことができるシステムも登場している。しかし、こうしたオープンドメインの対話システムにおいて、システムの応答は単純な一問一答にとどまっている。一方で、ユーザの複雑で曖昧な情報要求に対して、対象ドメインの知識を利用しながら、複数ターンにわたって対話を行うシステムも求められている。これは単純なキーワードベースの検索ではなく、観光地やレストランなどについてより詳細な情報の案内を行うものである。このようなアプリケーションは、対象とするドメインの知識を記述した文書を検索することによって実現することができる [Kawahara 09]。例えば、観光ガイドブックや Wikipedia 中の文書を利用して観光地のナビゲーションを行うシステム [Misu 10] が挙げられる。このような対象 (ドメイン) は多様にあるので、音声対話システムに必要な要素を対象ドメインの文書テキストから自動で構築できることが望まれる [吉野 11b]。

音声対話システムにおける音声認識モジュールは、ドメインと発話スタイルに適合した言語モデルを必要とする。既存の大語彙音声認識システムは、ドメイン特有の固有表現をカバーすることが難しいが、固有表現の認識誤りは、情報案内において致命的である。したがって、ドメイン毎に音声認識用の言語モデルを構築する必要があるが、そのための学習データが大量に用意できるという前提は現実的ではない。そこで、対象ドメインの文書を種として Web から関連した文章を収集する手法が検討されてきた [Sarikaya 05, Sethy 05, 翠 07, Bulyko 07]。これらは対象ドメインと話し言葉表現を間接的にカバーしようとするアプローチであるが、結果として多くの対象ドメイン以外の文を含んでしまうという問題点があった。これに対して、本論文では対象ドメインとの意味的な類似性に着目して、適合した文を選択する手法を提案する。

2. 提案手法の概要

提案手法の概要を図 1 に示す。本研究では、対話システムが対象ドメインの文書集合 D を検索して情報案内を行うことを想定する。また、言語モデル学習のために

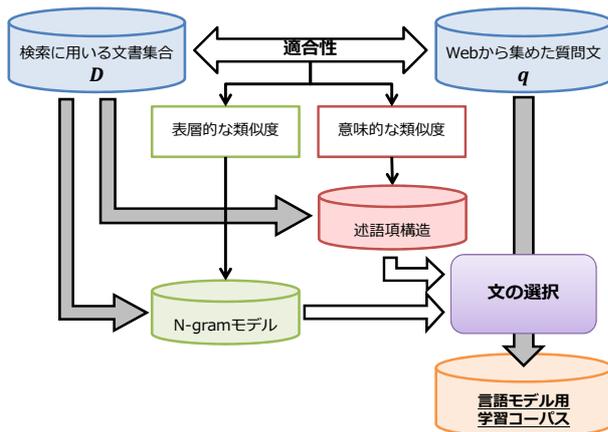


図1 提案手法の概要

Web から収集した文 q の集合を利用する．文書集合 D は書き言葉なので，言語モデルの学習データとして適当でなく，また Web から集めた文 q は対象ドメインに合致しないものが多い．従来手法では，ドメイン文書集合 D に対する q の単語系列の表層的な類似度を定義し，Web から収集した文の選択を行う．この手法について3章で述べる．本研究では，述語項構造に基づく意味情報を利用することによって，深層的な類似度を定義する．この新たな手法について4章で述べる．パープレキシティによる選択は生成モデルを仮定しているのに対して，提案手法では識別的なモデルによる文選択を行う．さらに，上記の2種類の文選択手法を併用する手法を検討する．これにより，2つの手法の異なる特性が効果的に働くことが期待できる．これについて5章で述べる．

3. N-gram モデルの表層的類似度

質問文 q が検索対象の文書集合 D にどの程度適合するかの表層的な指標として，N-gram モデルにおける KL 距離を導入する．KL 距離とは2つの確率分布の差を測る尺度であり [Kullback 51]，質問文 q と文書集合 D の KL 距離は以下のように定義される． w_i は質問文 q に含まれる単語である．

$$KL(q||D) = \sum_i P_q(w_i) \log_2 \frac{P_q(w_i)}{P_D(w_i)} \quad (1)$$

ここで P_D と P_q は D と q の言語モデルによって生成される確率であり，N-gram モデルによって与えられる確率で定義する．本研究では3-gram によるモデルを用いる． q に含まれる N-gram が一意である場合，すなわち同一の N-gram 連鎖が2回以上出現しない場合に $P_q(w_i) = 1$ となる． q は質問文一文を想定しており，言語モデルの確率を3-gram で与えると，多くの場合この条件が満たされる．そこで，式(1)は以下のように近似することが

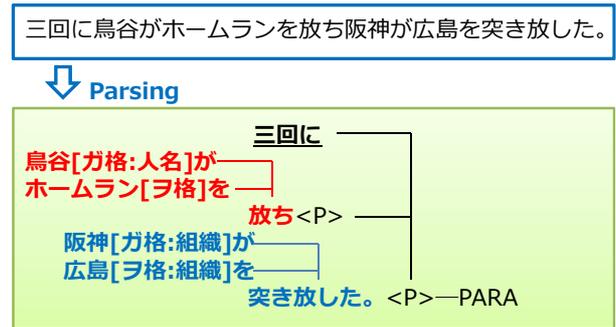


図2 述語項構造の例

できる．

$$KL(q||D) \approx \sum_i \log_2 \frac{1}{P_D(w_i)} \quad (2)$$

$$= - \sum_i \log_2 P_D(w_i). \quad (3)$$

Web テキストを利用する先行研究において，検索対象の文書集合 D の言語モデルによる質問文 q のテストセットパープレキシティを利用した文選択が行われていた [翠07] が，テストセットパープレキシティ $PP(q, D)$ は以下のように定義される．

$$H(q, D) = -\frac{1}{n} \sum_{i=1}^n \log_2 P_D(w_i). \quad (4)$$

$$PP(q, D) = 2^{H(q, D)}. \quad (5)$$

これは文書集合と質問文の KL 距離を測ることとほぼ等価であると解釈できる．

4. 述語項構造に基づく意味的類似度

3章で述べた文選択手法では，検索対象文書集合 D と質問文 q の表層的な類似度を利用していた．しかし，文の表層的な適合性のみでは，検索対象文書中にある構造的・意味的なレベルで合致した文を選択することは難しい．また，多くの対話システムでは意味的な情報を抽出することが求められている．そこで本論文では，述語項構造に着目した意味的類似度を定義し，深層的に合致する文を選択する手法を提案する．KL 距離あるいはパープレキシティに基づく手法では全ての単語の表層的な類似度を測っていたのに対して，本章では述語項構造における要素に着目した類似度の定義を行う．また，前章では N-gram モデルに基づく生成モデルを利用していたが，本章では識別モデルを導入する．

4.1 意味的類似度の定義

意味的類似度の定義を行うために，述語項構造に着目する．述語項構造は「誰が」「何を」などの格要素と，動詞や事態を表す名詞である述語を結びつけたものであ

る [Fillmore 68] . こうした構造は、古くから自然言語処理において利用されてきた . 述語項構造の抽出例を図 2 に示す . 本研究では、述語とそれに対する格要素及び格情報からなる述語項に着目する . 図 2 の例では「放つ」という述語に対して、格要素「鳥谷」が「ガ格」で、格要素「ホームラン」が「ヲ格」でかかっている . この最小単位である「鳥谷 (ガ格) - 打つ」、「ホームラン (ヲ格) - 打つ」が述語項である . 以降では、格情報を含めた述語を w_p , 格要素を w_a と表記する . 近年こうした構造 (格フレーム) の自動構築も盛んに行われており [Kawahara 06, Sasano 10] , パーザ JUMAN/KNP *1 では、自動構築された大規模格フレームに基づいて、省略された項の同定も含む述語項構造解析が行われる [笹野 13] .

しかし、文書集合を解析して得られる述語項構造全てが情報案内に有用なわけではなく、ドメインに依存した有用な情報構造のパターンがあることが知られている [Grishman 03, Ramshaw 05] . 例えば、野球ドメインにおいては「A (ガ格), B (二格) - 勝つ」「A (ガ格), B (ヲ格) - 打つ」といったパターンが重要であるが、経済ドメインでは「A (ガ格), B (ヲ格) - 売る」「A (ガ格), B (ヲ格) - 買収」などが重要なパターンとなる . こうしたドメイン依存の情報構造は、手動で定義することが一般的であったが [Grishman 03] , これを自動で抽出する手法が提案されている [Yoshino 11a] . このドメイン依存の情報構造を自動で定義する際には、ナイーブベイズ法を用いた手法が TF-IDF 法 よりも有効であることが報告されている .

そこで、ナイーブベイズ法によって定義されるドメインらしさを表す評価値を用いて意味的な類似度を定義する . 単語 w_i が与えられたとき、そのドメインが文書集合 D と一致する確率は、ベイズの定理を用いて以下のように定義できる .

$$P(D|w_i) = \frac{P(w_i|D) \times P(D)}{P(w_i)} . \quad (6)$$

これはディリクレスムージングによって以下のように近似される .

$$P(D|w_i) \simeq \frac{C(w_i, D) + P(D) \times \gamma}{C(w_i) + \gamma} . \quad (7)$$

ここで $C(w_i, D)$ は文書集合 D における単語 w_i の頻度であり、 $C(w_i)$ は D 以外も含む文書集合における単語 w_i の頻度である . γ は中華料理店過程を用いて推定されたディリクレ過程に基づくスムージング係数である [Teh 06] . 識別的なアプローチを取っているため、学習データとしてドメイン外データである \bar{D} が必要となるが、これに関しては文書集合 D と同じ出典から無作為に抽出したドメイン外コーパスを用いる . また、これにより $P(D)$ を推定する . 述語 ($w_{i,p}$) , 格要素 ($w_{i,a}$) から構成される

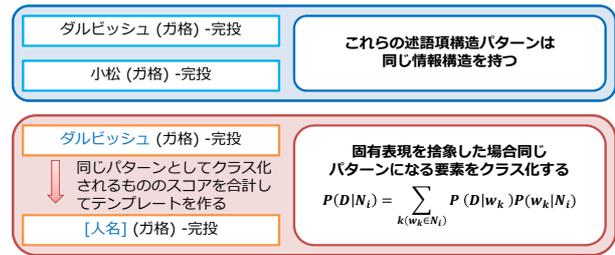


図 3 固有表現のクラス化

述語項 pa_i に対して、意味的類似度の評価値 $PA(D|pa_i)$ を以下のように両者に関する確率の幾何平均として定義する .

$$PA(D|pa_i) = \sqrt{P(D|w_{i,p}) \times P(D|w_{i,a})} . \quad (8)$$

4.2 固有表現のクラス化

統計的手法においては、データスパースネスの問題が学習データと評価データの不整合により生じ、特に固有表現において大きな問題となる . そこで、固有表現をクラス化して式 (8) の確率を計算することでこの問題の解決を図る . 固有表現は、図 2 に示されているように、人名や組織名などの固有名詞に自動でタグを付与したもので、述語項構造と同様に意味解析によって得られる . 今回は、固有表現を捨象した場合に同じパターンとなる述語項の確率を合計する . この例を図 3 に示す . この例では、「人名」クラスの格要素を持つ「ガ格」と述語「完投」が同一のパターンにまとめられている . 固有表現で捨象した場合に同じパターンとなる述語項を持つ確率の合計を、捨象された固有表現を持つ述語項に関する確率とする . 図 3 の例では「人名」クラスの格要素を持ち、格情報と述語が同一のパターンがクラス化されている . まとめられた固有表現のクラス N_i に対する確率は以下のように求める .

$$P(D|N_i) = \sum_{k(w_k \in N_i)} P(D|w_k)P(w_k|N_i) . \quad (9)$$

4.3 意味的類似度による文の選択

選択対象文 q 中に存在する述語項 pa_i 全てに関する $PA(D|pa_i)$ の平均をとり、 $PA(D|q)$ とする . この評価の例を図 4 に示す . この例では、入力文 q が 4 つの述語項 (「[人名] (ガ格) - 移籍:する」、「今:オフ (二格) - 行使:する」、「FA:権 (ヲ格) - 行使:する」、「他:球団 (二格) - 移籍:する」) を持っており、各述語項についての評価値 $PA(D|pa_i)$ が下線で示されている . これらの評価値の平均を計算することで、入力文 q の評価値を決定する . この評価値 $PA(D|q)$ が高いものを選択して言語モデルの学習データとする . これにより、検索対象の文書集合に、意味的に関連があるユーザ発話を認識しやすい言語モデルの構築が期待できる .

*1 <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>,
<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

述語項構造

$q =$ “イチローは今オフに F A 権を行使して他球団に移籍すると思いますか。”
 $pa =$ [“[人名]/カ格/移籍:する”, “今:オフ/二格/行使:する”,
 “F A:権/ヲ格/行使:する”, “他:球団/二格/移籍:する”]

各述語項の評価値

評価値	格要素	意味表現	述語
0.98201	フォーク	ヲ格	はじく:返す
0.98202	ホームラン	ヲ格	放つ
0.96353	F A:権	ヲ格	行使:する
0.95954	他:球団	二格	移籍:する
0.92919	今オフ	二格	行使:する
0.78062	[人名]	ガ格	放つ
0.68310	[人名]	ガ格	移籍:する
0.09994	株価	ガ格	下落:する
0.09994	負債	ガ格	拡大:する
	...		

図 4 意味的な類似度の例

5. 文選択手法の併用

3 章及び 4 章で異なる類似度を述べたが、これらを組み合わせ文選択を行うことも検討する。組み合わせの方法として、文の順位に基づく手法と文のスコアに基づく手法を検討する。

5.1 文の順位に基づく手法

3 章と 4 章で各文にドメイン文書集合 D に対する類似度を付与する手法を示したが、この評価値によって選択対象文 q を並び替え、順位 (PP_{rank} と PA_{rank}) を付与する。この順位の合計 ($PP_{rank} + PA_{rank}$) によって文を並び替え、文の選択に用いる。

5.2 文の評価値に基づく手法

3 章と 4 章で示した評価値を組み合わせ、新しい評価値を定義する。この際、2 つの評価値の値域を揃えるために、パープレキシティ PP を以下のシグモイド関数によって変換する。

$$PP'(q, D) = \frac{1}{1 + e^{-PP(q, D)}} \quad (10)$$

$PP'(q, D)$ と $PA(D|q)$ の混合比は試行の結果、3:7 と定めた。

6. 評価実験

前章までに述べた手法によって文選択を行い、選択された文から言語モデルを構築した上で、音声認識による評価を行った。また、音声認識結果からの意味理解についても評価した。

6.1 対象ドメインとシステムの構成

評価対象として、プロ野球ニュースに関する案内システム [吉野 11b] と京都観光に関する案内システム [Misu 10] でのユーザ発話音声を利用した。プロ野球案内システムには毎日新聞記事データベース (CD-毎日新聞データ

表 2 評価セットの詳細

タスク	話者数	発話数
京都観光案内	4	219
プロ野球案内	10	2,747

集 2000-2009) のうち「日本プロ野球」のタグが付与された記事を用い、京都観光案内システムには Wikipedia における京都関連文書を用いて、 $PP(q, D)$ と $PA(D|q)$ の統計量を学習した。音声認識用言語モデルの学習用テキストとしては Yahoo!知恵袋^{*2} から、質問タグが付与されているものを用いた。プロ野球案内システムには「エンターテインメント-野球」カテゴリのものを、京都観光案内システムは「旅行-国内」のカテゴリのものをそれぞれ用いた。この各質問文を $PP(q, D)$ と $PA(D|q)$ によって評価・並び替えし、それぞれのスコアによって選択された文から単語 3-gram モデルを学習した。選択する文数 (学習セット全体に対する割合) を変化させることにより、評価を行った。表 1 に学習セットの詳細を、表 2 に評価セットの詳細を示す。

6.2 音声認識精度による評価

音声認識用デコーダとしては Julius ver. 4.1.5^{*3} [Lee 01] を用いた。また、音響モデルには、文献 [鹿野 01] に付属する CD-ROM に含まれる性別非依存のトライフォンモデル (2000 状態 × 16 混合) を用いた。

音声認識の評価尺度として単語誤り率 (WER) を用いる。また、参考のために補正パープレキシティ (Adj. PP) を示す。補正パープレキシティでは学習テキスト全体から語彙を構築し、この中から、選択された学習テキストに出現しない単語の数によって未知語 (<UNK>) の確率を割る。これにより、語彙サイズが異なる言語モデル同士のテストセットパープレキシティの比較が可能になる。

プロ野球案内タスクの評価セットにおける単語誤り率を図 5 に、京都観光案内タスクの評価セットにおける単語誤り率を図 6 に示す。参考のためにそれぞれの補正パープレキシティを図 7、図 8 に示す。いずれも横軸は、学習に利用した質問文の割合である。PP が 3 章で述べた表層的な類似度 $PP(q, D)$ を用いて文選択を行った場合、PA が 4 章で述べた述語項構造に基づく類似度 $PA(D|q)$ を用いて文選択を行った場合、PP+PA が文の順位に基づく両手法の併用を行った場合である。5 章で述べた 2 つの併用手法については、グラフで表示すると違いがわからない程度であったので、順位に基づく手法 ($PP_{rank} + PA_{rank}$) のみを PP+PA として示している。

単語誤り率 (WER) においては、プロ野球案内タスク (図 5) で、提案手法である述語項構造に基づく類似度を利用した場合 (PA; text=7/10)、選択を行わない場合と比

*2 このコーパスは Yahoo!JAPAN と国立情報学研究所から提供を受けた。

*3 <http://julius.sourceforge.jp>

表1 学習セットの詳細

用途	タスク	コーパス名	文数
文選択器の学習 (D)	京都観光案内	Wikipedia	35,641
	プロ野球案内	毎日新聞データベース	176,852
音声認識用言語モデルの学習 (q)	京都観光案内	Yahoo!知恵袋コーパス:旅行-国内	679,588
	プロ野球案内	Yahoo!知恵袋コーパス:エンターテイメント-野球	403,602

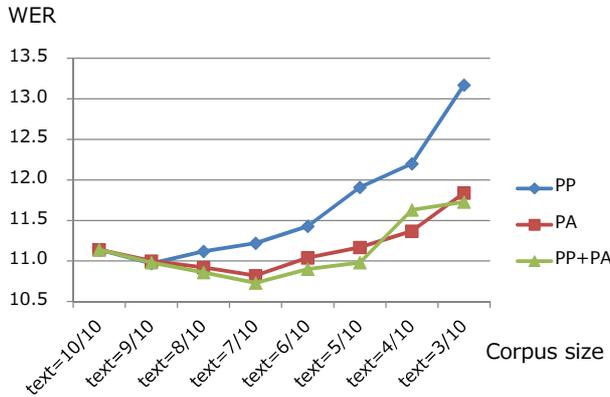


図5 プロ野球案内タスクにおける単語誤り率 (WER)

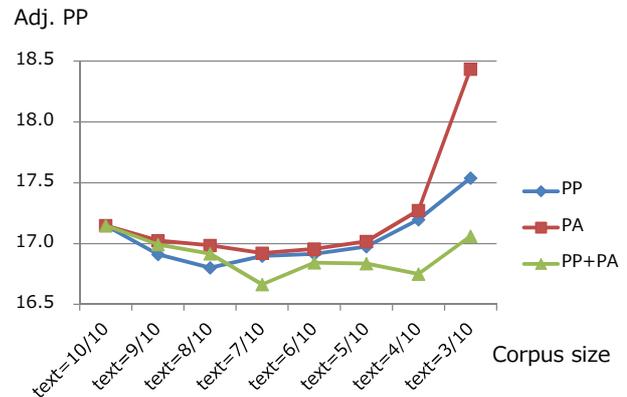


図7 プロ野球案内タスクにおける補正パープレキシティ (Adj. PP)

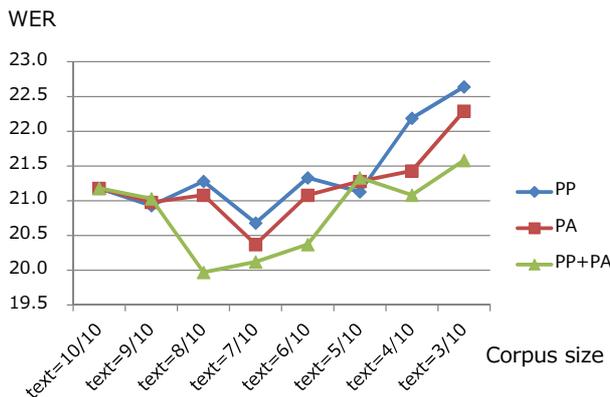


図6 京都観光案内タスクにおける単語誤り率 (WER)

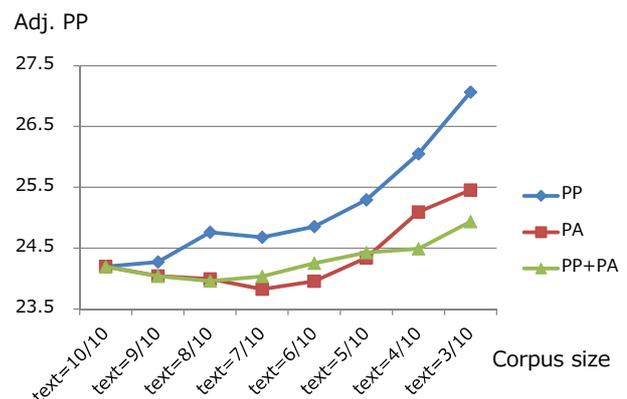


図8 京都観光案内タスクにおける補正パープレキシティ (Adj. PP)

較して有意な性能差が認められた (有意水準 $p < 0.05$) . ただしこの場合 (PA; text=7/10) と, 両手法を併用する場合 (PP+PA; text=7/10) では有意な差が見られなかった . 京都観光案内タスク (図6) においては, 併用する場合 (PP+PA) に全般的に改善が認められた .

文選択手法を実際のシステムに適用する場合, どのように選択する量 (図5や図6におけるグラフの点) を決定するかが大きな問題となる . 特に新しいタスクに適用する際に, 大規模な評価セットを用意しなくてもよいのが望ましい . 提案手法 (PA) では, 2つのタスク (図5, 図6) において text=7/10 が最適な性能になっているのに対して, 表層的な類似度による手法 (PP) では最適な点がかんり異なっている . 図6の京都観光案内タスクの評価セットを, 図5のプロ野球案内タスクの開発セットとみなすと, PP, PAともに text=7/10 の点が選択されるが, この場合に PA と PP の間に有意な性能差を認めること

ができる (有意水準 $p < 0.05$) . なお未知語率についてはプロットしていないが, 比較を行なっている text=7/10–text=10/10の間で PP と PA の未知語率の差は0.01% 未満であり, ほとんど影響はないと考えられる .

これらの結果から, 意味的な類似度を利用する提案手法により音声認識精度の向上が得られることが示された . また, いずれのドメインにおいても, 2つの類似度を併用した場合に最も良い性能が得られている .

6.3 意味理解による評価

次に, 音声対話における意味理解に関する評価を行った . 意味理解の評価には, 述語項誤り率 (PAER) を用いる . これは, 認識対象文中の述語項構造における「格要素/格情報/述語」の三つ組の正解精度で, 認識結果を述語項構造解析したものを正解と順番に対応づけて, 単語

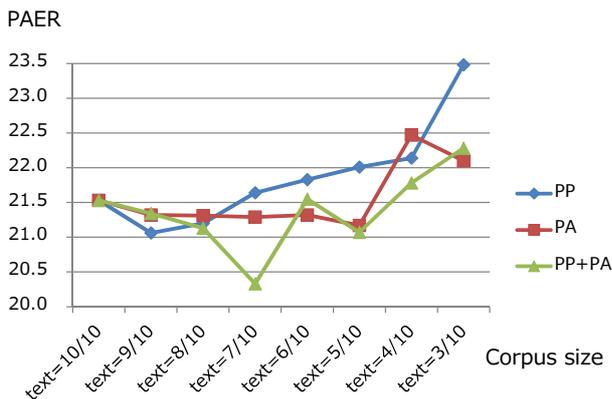


図9 プロ野球案内タスクにおける述語項誤り率 (PAER)

誤り率 (WER) と同様に計算する。この際、述語項の 3 つ組が全て正しく抽出できている場合に正解とする。京都観光案内タスクでは評価セットが小さく、述語項の数が少ないので、プロ野球案内タスクにおける述語項誤り率を図9に示す。2つの類似度を併用した場合 (PP+PA; text=7/10) に誤り率が最も低くなっており、選択を行わない場合の 21.5% から 20.4% まで改善している。述語項の数は多くない (2935 個) ので、統計的に有意な水準ではないが、深層的な情報の利用が意味理解の性能の改善に寄与していると考えられる。

7. 関連研究

Web から獲得した言語資源を音声認識の言語モデル構築に活用することは、Web の普及に伴って研究されてきた。例えば、Web から獲得した 3-gram の頻度を用いる手法 [Zhu 01] や、タスクごとに固有のキーワードを手動で設定して文を収集する手法 [Nishimura 01] が挙げられる。最も一般的な手法 [Sarikaya 05, Bulyko 07, Wan 06, Tsiartas 10] では、ドメインで特徴的な N -gram を Web の検索クエリとして文を収集する。こうした手法では種となる N -gram を獲得するためのコーパスが必要である。このような検索クエリを生成するために、対話システムの検索対象となる文書集合 [翠 07] や、講演におけるスライド [Munteanu 07, Kawahara 08]、初期的な音声認識結果 [Suzuki 06] などを用いることが検討されている。

一方で、Web から取得した文集合から言語モデル学習に適切な文を選択する手法も研究されてきた。選択のために最も一般的に用いられるのは、種として用いたテキストから構築した言語モデルによるパープレキシティである [Bulyko 07, 翠 07] が、BLEU スコアを用いる手法 [Sarikaya 05] や、トピックモデルの利用 [Sethy 05] なども検討されてきた。Masumura ら [Masumura 11] はナイーブベイズ法を用いた文選択を提案しているが、これらの先行研究における指標はいずれも文表層に関するもので、文の意味的な類似度まで考慮されていない。

8. ま と め

本論文では、音声対話システムのための言語モデル構築に利用する文の選択手法について提案した。既存手法である表層的なパープレキシティを用いた手法と比較して、述語項構造に基づく意味的な類似度を利用することで、音声対話システムのバックエンドに用いる知識ベースにより適合した文を選択することができる。この手法により、有意に音声認識精度を改善できることを示した。また、異なるドメインに適用することによって、手法の一般性を確認することができた。さらに音声対話システムにおける意味理解の精度が向上することも確認された。提案手法は、バックエンドに知識ベースや文書集合を持つような音声対話システムに対して、対話コーパスを一切収集することなく、ドメイン固有の言語モデルを構築することができるので、様々なドメインへの応用が期待される。

謝 辞

この研究は JSPS 特別研究員奨励費 254537 の助成を受けた。

◇ 参 考 文 献 ◇

- [Akbcak 05] Akbcak, M., Gao, Y., Gu, L., and Kuo, H.-K. J.: Rapid transition to new spoken dialogue domains: Language model training using knowledge from previous domain applications and web text resources, in *Proceedings of INTERSPEECH*, pp. 1873–1876 (2005)
- [Bulyko 07] Bulyko, I., Ostendorf, M., Siu, M., Ng, T., Stolcke, A., and Çetin, O.: Web resources for language modeling in conversational speech recognition, *ACM Trans. Speech Lang. Process.*, Vol. 5, No. 1, pp. 1:1–1:25 (2007)
- [Fillmore 68] Fillmore, C. J.: The case for case, in Bach, E. and Harms, R. eds., *Universals in Linguistic Theory* (1968)
- [Grishman 03] Grishman, R.: Discovery Methods for Information Extraction, in *Proceedings of ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pp. 243–247 (2003)
- [Hakkani-Tür 06] Hakkani-Tür, D. and Rahim, M.: Bootstrapping language models for spoken dialog systems from the world wide web, in *Proceedings of IEEE-ICASSP: IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 1, pp. 1065–1068 (2006)
- [Kawahara 06] Kawahara, D. and Kurohashi, S.: A Fully-Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis, in *Proceedings of HLT-NAACL: the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pp. 176–183 (2006)
- [Kawahara 08] Kawahara, T., Nemoto, Y., and Akita, Y.: Automatic lecture transcription by exploiting presentation slide information for language model adaptation, in *Proceedings of IEEE-ICASSP: IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4929–4932 (2008)
- [Kawahara 09] Kawahara, T.: New perspectives on spoken language understanding: Does machine need to fully understand speech?, in *Proceedings of ASRU: IEEE Workshop on Automatic Speech Recognition & Understanding*, pp. 46–50 (2009)
- [Kullback 51] Kullback, S. and Leibler, R. A.: On information and sufficiency, *The Annals of Mathematical Statistics*, Vol. 22, No. 1, pp. 79–86 (1951)
- [Lee 01] Lee, A., Kawahara, T., and Shikano, K.: Julius—an open source real-time large vocabulary recognition engine, in *Proceedings*

of *EUROSPEECH2001: the 7th European Conference on Speech Communication and Technology*, pp. 1691–1694 (2001)

- [Masumura 11] Masumura, R., Hahm, S., and Ito, A.: Training a language model using webdata for large vocabulary Japanese spontaneous speech recognition, in *Proceedings of INTERSPEECH*, pp. 1465–1468 (2011)
- [翠 07] 翠 輝久, 河原 達也: ドメインとスタイルを考慮した Web テキストの選択による音声対話システム用言語モデルの構築, 電子情報通信学会論文誌, Vol. J90-D, No. 11, pp. 3024–3032 (2007)
- [Misu 10] Misu, T. and Kawahara, T.: Bayes Risk-based Dialogue Management for Document Retrieval System with Speech Interface, *Speech Communication*, Vol. 52, No. 1, pp. 61–71 (2010)
- [Munteanu 07] Munteanu, C., Penn, G., and Baecker, R.: Web-based language modelling for automatic lecture transcription, in *Proceedings of INTERSPEECH*, pp. 2353–2356 (2007)
- [Nishimura 01] Nishimura, R., Komatsu, K., Kuroda, Y., Nagatomo, K., Lee, A., Saruwatari, H., and Shikano, K.: Automatic n-gram language model creation from web resources, in *Proceedings of EUROSPEECH2001: the 7th European Conference on Speech Communication and Technology*, pp. 5181–5184 (2001)
- [Ramshaw 05] Ramshaw, L. A. and Weischedel, R. M.: Information Extraction, in *Proceedings of IEEE-ICASSP: IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 969–972 (2005)
- [Sarikaya 05] Sarikaya, R., Gravano, A., and Gao, Y.: Rapid Language Model Development Using External Resources for New Spoken Dialog Domains, in *Proceedings of IEEE-ICASSP: IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 573–576 (2005)
- [Sasano 10] Sasano, R., Kawahara, D., and Kurohashi, S.: The Effect of Corpus Size on Case Frame Acquisition for Predicate-Argument Structure Analysis, *IEICE Transactions*, Vol. 93-D, No. 6, pp. 1361–1368 (2010)
- [笹野 13] 笹野 遼平, 河原 大輔, 黒橋 禎夫, 奥村 学: 構文・述語項構造解析システム KNP の解析の流れと特徴, 言語処理学会第 19 回年次大会 発表論文集, pp. 110–113 (2013)
- [Sethy 05] Sethy, A., Georgiou, P. G., and Narayanan, S.: Building Topic Specific Language Models from Webdata Using Competitive Models, in *Proceedings of INTERSPEECH*, pp. 1293–1296 (2005)
- [鹿野 01] 鹿野 清宏, 伊藤 克亘, 河原 達也, 武田 一哉, 山本 幹雄: 音声認識システム, オーム社 (2001)
- [Suzuki 06] Suzuki, M., Kajiuira, Y., Ito, A., and Makino, S.: Unsupervised language model adaptation based on automatic text collection from WWW, in *Proceedings of INTERSPEECH*, pp. 2202–2205 (2006)
- [Teh 06] Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M.: Hierarchical Dirichlet Processes, *Journal of the American Statistical Association*, Vol. 101, pp. 1566–1581 (2006)
- [Tsiartas 10] Tsiartas, A., Georgiou, P., and Narayanan, S.: Language model adaptation using www documents obtained by utterance-based queries, in *Proceedings of IEEE-ICASSP: IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 5406–5409 (2010)
- [Wan 06] Wan, V. and Hain, T.: Strategies for language model webdata collection, in *Proceedings of IEEE-ICASSP: IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 1069–1072 (2006)
- [Yoshino 11a] Yoshino, K., Mori, S., and Kawahara, T.: Spoken Dialogue System based on Information Extraction using Similarity of Predicate Argument Structures, in *Proceedings of the SIGDIAL: the 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 59–66 (2011)
- [吉野 11b] 吉野 幸一郎, 森 信介, 河原 達也: 述語項の類似度に基づく情報抽出・推薦を行う音声対話システム, 情報処理学会論文誌, Vol. 52, No. 12, pp. 3386–3397 (2011)
- [Zhu 01] Zhu, X. and Rosenfeld, R.: Improving trigram language modeling with the world wide web, in *Proceedings of IEEE-ICASSP: IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 533–536 (2001)

[担当委員: 南 泰裕]

2013 年 4 月 9 日 受理

著者紹介



吉野 幸一郎

2009 年 慶応義塾大学環境情報学部卒業。2011 年 京都大学大学院情報科学研究科修士課程修了。同年 同大学院博士後期課程に進学。2013 年より日本学術振興会特別研究員 (DC2)。音声言語処理及び自然言語処理, 特に音声対話システムに関する研究に従事。情報処理学会, 言語処理学会各会員。



森 信介

1993 年 京都大学工学部電気電子工学科卒業。1995 年 同大学院修士課程修了。1998 年 同博士後期課程修了。同年 日本アイ・ピー・エム (株) 入社。2007 年 京都大学学術情報メディアセンター准教授。京都大学博士 (工学)。1997 年 情報処理学会山下記念研究賞受賞。2010 年 情報処理学会論文賞受賞。2010 年 第 58 回電気科学技術奨励賞受賞。情報処理学会, 言語処理学会各会員。



河原 達也(正会員)

1987 年 京都大学工学部情報工学科卒業。1989 年 同大学院修士課程修了。1990 年 同博士後期課程退学。同年 京都大学工学部助手。1995 年 同助教授。1998 年 同大学情報科学研究科助教授。2003 年 同大学学術情報メディアセンター教授。現在に至る。この間, 1995 年～1996 年 米国・ベル研究所客員研究員。1998 年～2006 年 ATR 客員研究員。1999 年～2004 年 国立国語研究所非常勤研究員。2006 年～情報通信研究機構短時間研究員・招へい専門員。音声言語処理, 特に音声認識及び対話システムに関する研究に従事。京都大学博士 (工学)。科学技術分野の文部科学大臣表彰 (2012 年度), 日本音響学会から粟屋潔学術奨励賞 (1997 年度), 情報処理学会から坂井記念特別賞 (2000 年度), 喜安記念業績賞 (2011 年度), 論文賞 (2012 年度) を受賞。IEEE SPS Speech TC 委員, IEEE ASRU 2007 General Chair, INTERSPEECH 2010 Tutorial Chair, IEEE ICASSP 2012 Local Arrangement Chair, 言語処理学会理事, 情報処理学会音声言語情報処理研究会主査を歴任。情報処理学会, 日本音響学会, 電子情報通信学会, 言語処理学会, IEEE 各会員。