Vol.2012-NL-206 No.3
Vol.2012-SLP-91 No.3
2012/5/10

述語項構造を介した文の変換と選択に基づく音声対話用言語モデルの構築

吉野 幸一郎[1]　森 信介[1]　河原 達也[1]

概要：文書集合をバックエンドとして情報検索・案内を行う音声対話システムのための、言語モデル構築手法について述べる。提案手法では、検索対象である文書集合中の書き言葉のテキストに対して、音声対話で想定される話し言葉の質問文への変換を行う。この変換では、係り受け解析に基づいて、述語に係る文節と述語の変換により自然な質問文を生成する。次に、検索対象となるドメインにおいて有用な情報を持ち、検索される可能性が高い文の選択を行う。この選択のために、当該文書集合における統計量に基づいて述語項構造テンプレートを定義する。構築した言語モデルを音声認識で評価した結果、提案手法の有効性、特に文選択の効果を確認した。

キーワード：音声対話システム, 言語モデル, 述語項構造

Language Modeling for Spoken Dialogue System based on Sentence Transformation and Filtering using Predicate-Argument Structures

Koichiro Yoshino[1]　Shinsuke Mori[1]　Tatsuya Kawahara[1]

Abstract: We present a novel scheme of language modeling for a spoken dialogue system by effectively exploiting the back-end documents the system uses for information navigation. The proposed method first converts sentences in the document, which are written and plain style, into spoken question-style queries, which are expected in spoken dialogue. In this process, we conduct dependency analysis to extract verbs and relevant phrases to generate natural sentences by applying transformation rules. Then, we select sentences which have useful information relevant to the target domain and thus are more likely to be queried. For this purpose, we define predicate-argument (P-A) templates based on a statistical measure in the target document. An experimental evaluation shows that the proposed method outperforms the conventional method in ASR performance, and the sentence selection based on the P-A templates is effective.

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

1. はじめに

対話を行うことによって、ユーザの要求や嗜好に応じた情報案内を行う音声対話システムが求められている。これは従来の音声検索システムが行なっているような単純なキーワードの一致によるものではなく、観光案内やニュース情報の案内など、ユーザの曖昧で複雑な情報要求にも応えるものである。このようなタスクでは、「おすすめの場所はどこですか」などの漠然とした質問が発せられても、システムは最も関連のある情報を提示する必要がある。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。また、このような音声対話システムを実現するためには、ドメインを限定し、当該ドメインの知識を利用する必要がある [1]。我々はこのような枠組みに基づいて、Wikipedia 中の文書を利用し観光案内をするシステム [2] や、インタラクティブにニュース検索対話を行うシステム

[1] 京都市左京区吉田本町 京都大学 情報学研究科
School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, Japan

© 2012 Information Processing Society of Japan

を提案している [3]。

音声対話システムにおける音声認識モジュールは、タスクやドメイン、そしてスタイルに適応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などの問題があり、固有表現の認識誤りは、情報検索には致命的である。また、音声認識用言語モデルは大規模な認識対象文から構築する必要があるが、認識対象文が大量に用意できるという設定は現実的ではない。そこで次の 2 つの適応手法が、音声対話用音声認識において一般に利用されてきた。1 つは、対象ドメインの文書のテキストと、話し言葉表現を含んだ対話コーパスを混合する手法である [4]。もう 1 つは、Web から関連した文章を収集し、話し言葉表現を含むコーパスを構築する手法である [5], [6], [7], [8]。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、結果として多くの認識対象以外の文を含んでしまうという問題点があった。

本研究では、より直接的に書き言葉の検索対象文書テキストから、話し言葉形式の質問文を生成するアプローチを考える。音声対話システムにおいて想定される質問文を生成することができれば、高い認識率を期待できる [9]。しかし、文書中の全ての文が質問文生成に適切な表現や情報を含んでいるわけではない。重要な情報構造はドメインに依存しており、適切な情報を抽出して利用する必要がある [10]。こうした情報抽出のためのテンプレートは、人手で作ることが一般的であったが、Web に存在する様々なドメインに対して適用するには非常にコストが高く、現実的ではない。我々は、先行研究において、ドメインに依存した情報抽出のためのテンプレートを自動で抽出し、それを情報提示に利用するシステムを構築した [11]。

本稿ではこの枠組みを拡張して、音声対話用言語モデルの構築を行う。提案手法は 2 つの処理からなる。最初に文の係り受け構造、述語項構造を利用した文の変換を行い、次にドメインに依存した情報抽出テンプレートを用いて文を選択する。この手法により構築した音声対話用言語モデルの評価のため、プロ野球ニュースに関する情報案内システム [11] の検索対象である文書集合 (=新聞記事データベース) からモデルの構築を行い、その認識率を確認した。

2. 提案手法の概要

提案手法の概要を**図 1** に示す。最初に、新聞記事データベースを JUMAN[*1] と KNP[*2] で解析し、係り受け構造と述語項構造を取得する。まず、係り受け構造上の述語に着目する。述語項構造とは、文中の述語とそれに対する格要素と意味表現のペアからなるものである。これらの構造と述語に対する単純なルールを用いて、解析結果から質問文形式への変換を行う。次に、ドメインに依存した述語項構造テンプレートを、述語項構造解析結果から学習する。このテンプレートを用いて質問文の選択を行い、情報案内のために必要な語彙と表現を含んだ言語モデルを学習する。

[*1] http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman.html
[*2] http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp.html

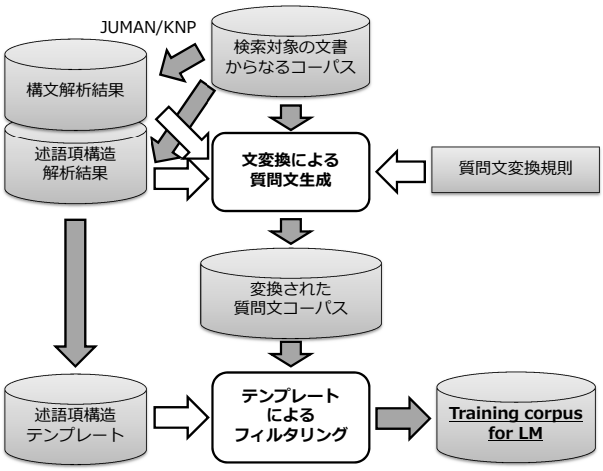


図 1 提案手法の概要

3. 文構造を利用した文変換

係り受け構造と述語項構造を利用した文変換について述べる。音声対話による情報案内システムは、その音声認識モジュールにおいて 2 つの問題を抱えている。1 つは、書き言葉の検索対象文書と話し言葉によるユーザ質問の形式が異なることである。日本語や英語においては、これらの主な差違は動詞や述語に表れる。そこで、動詞を特定し、文を述語項構造の構成要素ごとに分解し、述語に対して質問文形式への変換を行うルールを適用する。もう 1 つの問題は、検索対象の文書中には情報検索に冗長なものが多く含まれるという点である。そこで述語に係る文の係り受け構造に着目し、不要な文節を削除することで冗長性を削減する。

この処理の例を**図 2** に示す。この例では、2 つの述語「放つ」と「突き放す」が変換の対象となり、それらの述語を根とする部分木が質問文形式への変換に用いられる。

3.1 ルールによる述語の変換

まず、文中の述語を同定する。述語は用言と事態性名詞の 2 種類に分類することができ、用言には動詞、形容詞、形容動詞（ナ形容詞）が分類される。事態性名詞とは、文中において事態を表す名詞であり [12], [13]、名詞でありながら用言と同様の意味役割を持つ。**表 1** に各述語ごとの変換規則を示す。述語ごとに複数の変換規則が存在するが、全ての変換規則を適用し、複数パターンの文を生成する。

3.2 文生成

規則による変換を行った後、変換された述語を根とする係り受け木上の部分木を利用して質問文生成を行う。図 2

© 2012 Information Processing Society of Japan

表 1 述語の変換規則

述語タイプ	品詞	規則
用言	動詞&動詞性接尾辞	**連用形** + (/**ます** + **か**｜**まし** + **たか**/)
	イ形容詞（形容詞）&イ形容詞性接尾辞	**基本形** + (/**です** + **か**｜**でし** + **たか**/)
	ナ形容詞（形容動詞）&ナ形容詞性接尾辞	**語幹** + (/**です** + **か**｜**でし** + **たか**/)
事態性名詞	一般名詞&名詞性接尾辞	**原形** + (/**です** + **か**｜**でし** + **たか**/)
	サ変名詞&サ変名詞性接尾辞	**原形** +(/**し** + **ます** + **か**｜**し** + **まし** + **たか**/)

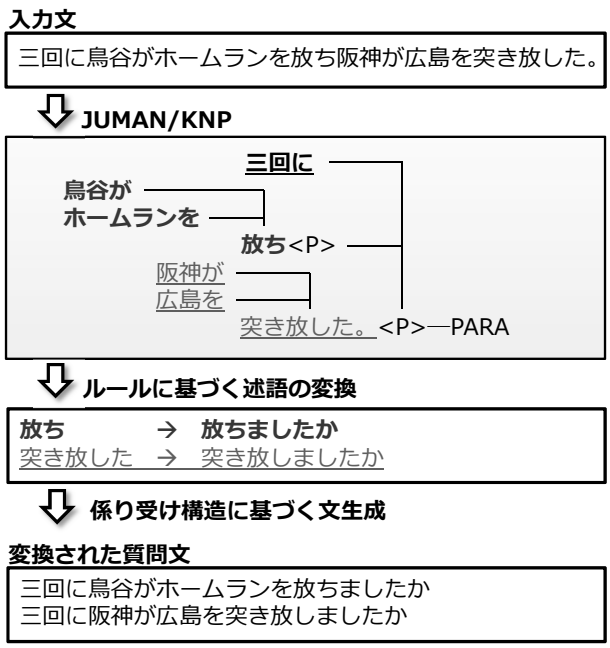


図 2 書き言葉から話し言葉への文変換の例

に挙げた例では、「放つ」の部分木は「鳥谷が」「ホームランを」「三回に」の文節になる。これらを元の文における出現順序を保持したまま、述語と合わせて質問文として生成する。この例の中では「三回に」は「放つ」と「突き放す」に並列して係っているが、こうした場合はいずれの述語の文生成にも利用される。

本研究では、述語に対する単純な変換規則によってYes/No 型の質問を生成したが、これらの構造情報を利用することで WH タイプの質問文を生成することもできる。例えば、格要素に付与される固有表現タグ「人名」を参照することで、Who タイプの質問文を生成することができる。

4. 述語項構造テンプレートを用いた文の選択

述語項構造テンプレートを定義し、検索対象のドメインにおいて有用な情報構造を抽出する。このテンプレートによって、前節で生成された質問文を選択する。

4.1 ドメイン固有の述語項構造テンプレートの抽出

述語項構造は、意味解析によって自動で得られる情報構造である。しかし、全ての述語項構造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。例えば、野球ドメインにおいて重要なパターンは「A（ガ格）, B（ニ格）-勝つ」「A（ガ格）,B（ヲ格）-打つ」といったパターンであるが、経済ドメインでは「A（ガ格）, B（ヲ格）-売る」「A（ガ格）, B（ヲ格）-買収」などのパターンが重要になってくる。我々はこうしたドメイン依存の述語項構造テンプレートを自動で定義する手法を提案している [11]。

先行研究において Naive Bayes 法を用いた手法が有効であることを示しており、これを利用する。この手法では、単語 w_i が与えられたときそのドメイン t である確率を次のように定義する。

$$P(t|w_i) = \frac{C(w_i,t) + D_t\gamma}{C(w_i) + \gamma}. \quad (1)$$

ここで、γ は Chinese Restaurant Processes を用いて推定されたディリクレ過程によるスムージング係数である [14]。D_t はコーパスサイズによるドメイン t の正規化係数であり、次のように定義する。

$$D_t = \frac{\sum_j C(w_j,t)}{\sum_k C(w_k)}. \quad (2)$$

述語項構造テンプレートの評価値は、述語項構造内の述語、格要素、意味表現である各単語の $P(t|w_i)$ の平均から求められる。ここで、2 通りの計算手法を試行する。1 つは述語と意味表現を 1 つの組として式 (1) の計算を行うものであり、もう 1 つは格要素と意味表現を 1 つの組として計算を行うものである。これらは以下の式で与えられる。

$$\begin{cases} NB_{ps_a}(t|P\text{-}A) = \sqrt{P(t|w_{ps}) \times P(t|w_a)} \\ NB_{p_sa}(t|P\text{-}A) = \sqrt{P(t|w_p) \times P(t|w_{sa})} \end{cases} \quad (3)$$

統計的手法においては、しばしば固有表現のスパースネスの問題が学習セットとテストセットの不整合として現れてくる。そこで本研究においては、固有表現をクラス化して述語項構造テンプレートの評価値を計算することで、この問題を解決した。

4.2 述語項構造テンプレートによるフィルタリング

3 章の手法で生成された全ての質問文に対して、文中に存在する述語項構造全てから次に定義する (NB_s) の値を計算し、質問文の評価値とする。

$$NB_s = \frac{\sum_{i=1}^{n} NB(t|P\text{-}A)}{n}. \quad (4)$$

© 2012 Information Processing Society of Japan

述語項構造

s = “三回に鳥谷がホームランを放ち阪神が広島を突き放した”
P-A = [“[人名]/ガ格/放つ”,
“ホームラン/ヲ格/放つ”,
“[組織名]/ガ格/突き放す”,
“[組織名]/ヲ格/突き放す”]

述語項構造テンプレート

スコア	格要素	意味表現	述語
0.99519	リリーフ:陣	ガ格	踏ん張れる:ぬ
0.98201	フォーク	ヲ格	はじく:返す
0.98202	ホームラン	ヲ格	放つ
0.78062	[人名]	ガ格	放つ
0.76810	[組織名]	ガ格	突き放す
0.76780	[組織名]	ヲ格	突き放す
0.09994	株価	ガ格	下落:する
0.09994	負債	ガ格	拡大:する
	…		

述語項構造テンプレートによる文のスコア

NB_s = (0.78062+0.98202+0.76810+0.76780) / 4 = 0.824635

図 3 述語項構造テンプレートによる評価値の計算例

この評価の例を**図 3** に示す。例では入力文 s は 4 つの述語項構造を持っており、各構造について述語項構造テンプレートによって評価値を与える。これらの平均を計算することで、質問文の評価値を決定する。

各質問文は NB_s の値によってソートされ、値が高いものから言語モデルの学習に利用する。この手法によって、当該ドメインの対話においてユーザに発話されやすい文を選択することができ、より適合した言語モデルを構築することができる。

5. 評価実験

毎日新聞記事データベース (CD-毎日新聞データ集 2000–2009) のうち、日本プロ野球ドメインの記事 176,852 文を解析した結果、500,523 個の述語 (342,322 個の用言と 158,201 個の事態性名詞) が抽出され、そこから提案手法による言語モデルの学習を行った。構築された言語モデルの評価を行うため、201 個の質問文をテストセットとして用いた。このテストセットに対するパープレキシティと単語誤り率 (WER) を用いて評価を行った。語彙サイズの異なる言語モデルを比較するために補正パープレキシティを用い、補正パープレキシティを定義するための認識語彙サイズは、カットオフ 5 の条件下で学習コーパス全体から求めた。

比較のため、話し言葉コーパスを混合する手法を用いる。先述の新聞記事コーパスに加えて、Yahoo!知恵袋コーパス[*3]中に含まれる、野球ドメインの質問文 481,243 文を用いた。

文選択の効果を確認するために、ソートした文の上位から 50%、25%、12.5%、6.25%をそれぞれ言語モデルに利用した。このパープレキシティを**図 4** に、単語誤り率 (WER) を**図 5** に示す。PS_A と P_SA という 2 通りの評価スコアの計算方法を試行し、それらの効果について調べた。横軸は学習に用いたデータの量である。提案手法は、

[*3] 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。

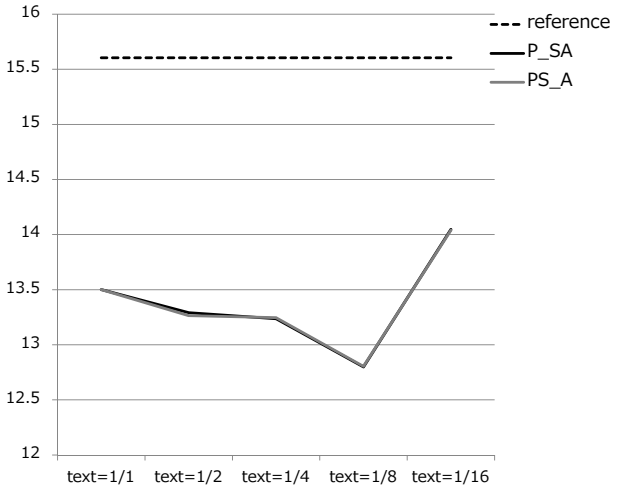


図 4 テストセットパープレキシティ

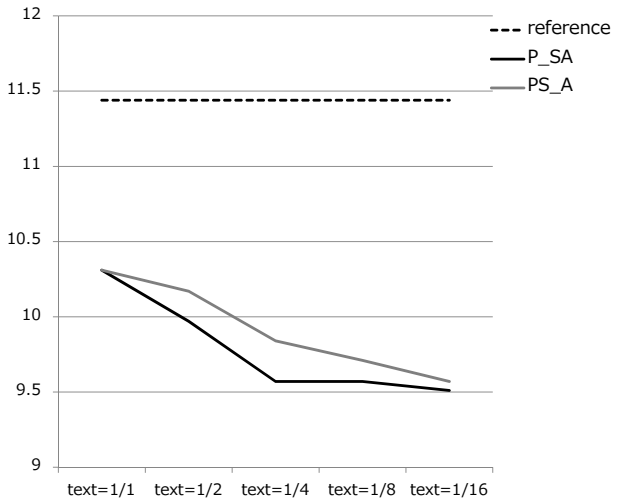


図 5 音声認識精度（単語誤り率）

従来手法と比較して、最大で単語誤り率 (WER) を 16.9%、パープレキシティを 18.0% 削減している。文選択を行わない場合 (text=1/1) と提案手法による文選択を行った場合を比較すると、パープレキシティで 5.2% 、単語誤り率で 7.8% の削減が確認された。これにより、提案手法がドメインを限定した情報案内システムにおける音声認識用言語モデルの構築手法として有効であることが示された。

6. まとめ

文の係り受け構造・述語項構造を用いて文変換と文選択を行うことにより、情報案内システムのための音声対話用言語モデルを構築する手法を提案した。この手法により、話し言葉コーパスなしに、検索対象の文書のみから既存手法を上回る認識精度を実現することができた。また、述語項構造テンプレートを利用した文の選択により、より情報案内システムに適合した言語モデルを構築することができた。今後は他のドメインや、さらに多量のデータにこの手法を適用することを考えている。

© 2012 Information Processing Society of Japan

参考文献

[1] Kawahara, T.: New perspectives on spoken language understanding: Does machine need to fully understand speech?, *Proc. IEEE-ASRU*, pp. 46–50 (2009).
[2] 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).
[3] 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).
[4] 駒谷和範，河原達也，清田陽司，黒橋禎夫，Fung, P.：柔軟な言語モデルとマッチングを用いた音声によるレストラン検索システム，情報処理学会研究報告，SLP-39-30, pp. 67–72 (2001).
[5] Sarikaya, R., Gravano, A. and Gao, Y.: Rapid Language Model Development Using External Resources for New Spoken Dialog Domains, *Proc. ICASSP*, Vol. 1, pp. 573–576 (2005).
[6] Sethy, A., Georgiou, P. G. and Narayanan, S.: Building Topic Specific Language Models from Webdata Using Competitive Models, *Proc. Interspeech*, pp. 1293–1296 (2005).
[7] Misu, T. and Kawahara, T.: A bootstrapping approach for developing language model of new spoken dialogue system by selecting Web texts, *INTERSPEECH*, pp. 9–13 (2006).
[8] 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).
[9] Istvan, V., Otake, K., Torisawa, K., Saeger, S. D., Misu, T., Matsuda, S. and Kazama, J.: Similarity Based Language Model Construction for Voice Activated Open-Domain Question Answering, *Proc. IJCNLP2011* (2011).
[10] Grishman, R.: Discovery Methods for Information Extraction, *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pp. 243–247 (2003).
[11] Yoshino, K., Mori, S. and Kawahara, T.: Spoken Dialogue System based on Information Extraction using Similarity of Predicate Argument Structures, *Proc. of SIGDIAL*, pp. 59–66 (2011).
[12] Grimshaw, J.: Argument Structure, *MIT Press* (1990).
[13] Komachi, M., Iida, R., Inui, K. and Matsumoto, Y.: Learning Based Argument Structure Analysis of Eventnouns in Japanese, *Proc. of the PACLING*, pp. 120–128 (2007).
[14] 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).

© 2012 Information Processing Society of Japan