

音声対話システム用クラス N-gram モデルによる ドメイン固有語の認識率向上

森 祥二郎¹ 駒谷 和範¹ 佐藤 理史¹

概要: データベース検索型音声対話システムでは、地名等のドメイン固有語の認識が重要である。本研究では、これをクラスとしたクラス N-gram モデルの構築を行う。これにはコーパスへのクラス付与が不可欠であり、ドメイン固有語を正しく認識するためには、ドメイン固有語が出現する文脈として適切な箇所にクラスを付与する必要がある。本研究では、機械学習を用いて適切な文脈にクラスを付与する。まず、データベース中のドメイン固有語とコーパスを組み合わせ、学習データを作成する。このデータからドメイン固有語が出現する文脈を学習してコーパスへクラスを付与し、これに基づきクラス N-gram モデルを構築する。評価実験ではドメイン固有語とコーパスの文字列マッチングによって作成したクラス N-gram モデルと比較し、ドメイン固有語の認識率向上を示す。

キーワード: 音声対話システム, 言語モデル, クラス N-gram モデル, 機械学習

Improving ASR Accuracy of Domain-Specific Words by Class N-gram Model for Spoken Dialogue Systems

SHOJIRO MORI¹ KAZUNORI KOMATANI¹ SATOSHI SATO¹

Abstract: Domain-specific words such as proper nouns must be correctly recognized in spoken dialogue systems. We construct a class N-gram model to recognize user utterances containing such domain-specific words. In order to recognize domain-specific words, it is necessary to assign semantic classes to appropriate context in a large-scale corpus. We assign semantic classes to the context by means of machine learning. First, we make training data by combining domain-specific words in target relational database for retrieval with a corpus. Next, we construct a class N-gram model from a large-scale corpus to which semantic classes are assigned by machine learning. The experimental evaluation showed that our proposed framework improves ASR accuracy of domain-specific words.

Keywords: spoken dialogue system, language model, class N-gram model, machine learning

1. はじめに

データベース検索型の音声対話システムでは、タスク遂行に必須であることから、ドメイン固有語の認識が重要である。ここでのドメイン固有語とは、検索対象データベースに含まれる、地名等のそのドメインに固有な名詞を指す。例えばドメインがある地方のレストラン検索の場合だと、店舗名や駅名などがこれにあたる。ここでドメインとは、

検索対象の個々のデータベースに対応する範囲とする。

音声対話システムにおける音声認識用の言語モデルには、次の3つが求められる。

- (1) 多様な発話パターンを認識できること
- (2) 当該ドメインのドメイン固有語が認識できること
- (3) 当該ドメインのコーパスが大量にはない状況でも構築できること

ユーザの発話パターンは多様であり、これらに対する柔軟性が必要である。ドメイン固有語は、データベース検索に必須な情報であるため、認識できなければならない。この

¹ 名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University

ような言語モデルは、当該ドメインの大規模コーパスから学習できるのが望ましいが、新たなドメインでの音声対話システム構築時に、そのドメインの大規模コーパスが存在すると仮定するのは現実的でない。

本研究では、システム構築時に利用できる、当該ドメインの小規模コーパスと類似ドメインの大規模コーパスを併用して、クラス N-gram モデルを構築する。当該ドメインのコーパスは、数百文程度ならば人手で用意できるとする。類似ドメインの大規模コーパスは、Web などから取得可能であるとする。クラスとクラス内単語は検索対象データベースに基づき決定する。クラス内単語をドメイン固有語とすることで、これを認識できる言語モデルの構築を目指す。ドメイン固有語をクラス内単語とした研究には [1] がある。文献 [1] では、当該ドメインの小規模コーパスに出現するドメイン固有語にクラスを付与し、このクラス付きコーパスと類似ドメイン大規模コーパスとを混合して、当該ドメインのクラス N-gram モデルを構築している。これに対して本研究では、類似ドメイン大規模コーパスにもクラスを付与する。我々は先に、類似ドメイン大規模コーパスにもクラスを付与することで、ドメイン固有語の音声認識率が向上することを確認した [2]。本稿でも同様に、類似ドメイン大規模コーパスにもクラスを付与し、ドメイン固有語の認識性能の向上を図る。

本研究では、ドメイン固有語が出現する文脈に基づいたクラス付与を行う。類似ドメイン大規模コーパスへのクラス付与は、クラスを認識しやすくするため、ドメイン固有語が出現する文脈として適切な箇所に、可能な限り多く付与することが望ましい。このため、クラス付与対象の単語がドメイン固有語として不適切であっても、適切な箇所に出現するのであればクラスを付与する。コーパス中の適切な箇所に、可能な限りクラスを付与し、クラス N-gram モデルを構築することで、ドメイン固有語の音声認識率の向上を図る。先の我々の研究では、検索対象データベースには存在しないが、適切な箇所に出現する単語を抽出し、コーパス中の全てのその単語にクラスを付与することで、クラス付与箇所の増加を図った [2]。この手法では、抽出した単語を用いてクラス付与を行うと、不適切な箇所にもクラスを付与するという問題があった。これは、抽出した箇所では適切であったとしても、別の箇所で出現する場合にも適切であるとは限らないためである。

本稿では、大規模コーパスへのクラス付与手法と、それに基づき構築したクラス N-gram モデルの音声認識性能について示す。まず 2 章で関連研究について述べる。次に 3 章で、本研究での問題設定と適切なクラス付与について述べる。4 章では、これを実現するための提案手法の枠組みを述べる。5 章で構築したクラス N-gram モデルの評価を行い、提案する枠組みの性能やその上限について議論する。6 章は本稿のまとめである。

2. 関連研究

クラス N-gram モデルに関する先行研究として [3] [4] などがある。これらは、パープレキシティの向上や小規模コーパスからの言語モデル構築を目的とした、クラスタリングの研究である。これらに対し、本研究は検索対象データベース中のドメイン固有語を、大規模コーパスから作成する言語モデルに組み込むことが目的である点で異なる。

言語モデル構築に利用する、類似ドメイン大規模コーパスの収集に関する研究としては、[5] [6] などがある。これらは、Web などからテキストを収集している。当該ドメインの小規模な言語モデルを用いて、収集したテキストから当該ドメインと類似した内容の文を選択している。類似度には、単語パープレキシティや相対エントロピーを用いている。本研究は、これらの手法で収集できるコーパスを用いてクラス N-gram モデルを構築するという位置付けになる。

クラス付与に関する研究としては、正解クラスが付与されたコーパスが大量に利用できる場合には、これを学習データとし、未知のドメイン固有語 (固有表現) を抽出する手法が提案されている [7,8]。これらの研究では、単語に人名や組織名などのラベル (クラス) を付与したコーパスを学習データとし、機械学習を行う。そして学習したモデルを用いて、抽出対象となるコーパスにラベルを付与し、ラベルが付与された単語を抽出する。このとき、学習データには存在しない未知の固有名詞も抽出対象コーパスから抽出する。これらの研究では、機械学習に用いる大規模なラベル付きコーパスが事前に用意されているという前提で行われている。これに対して、本研究が想定する状況では、大量のラベル付き学習データは事前に用意されていない。

そこで本研究では、まず機械学習のための学習データの作成から行う。前提において利用可能な資源から、学習データを作成する。次に、作成した学習データを用いて機械学習を行い、それに基づきコーパスにクラスを付与する。

3. 文脈に基づくドメイン固有語へのクラス付与

3.1 問題設定

本研究では、大規模コーパスにクラス付与を行うことで、クラス付き大規模コーパスを作成し、クラス N-gram モデルを構築する。クラス付き大規模コーパスの作成時には、音声対話システム構築時に用意される次の 3 つの資源が利用可能であるとする。

(1) 当該ドメインの検索対象データベース

システムの検索対象となるデータベースであり、システムが認識すべきドメイン固有語が登録されている。登録されているドメイン固有語の種類は数百から数千程度である。

- (1) ... のおすすめ のそば/FOOD を教えて ...
- (2) ... 駅のそば/FOOD があるクッチ・ナカサイ という ...
- (3) ... おすすめ の レシピ があれば教えて ...

図1 類似ドメインのコーパスとデータベースの一致をとってクラスを付与した例(レストラン検索ドメインの場合)

(2) 当該ドメイン小規模コーパス

システム構築時にシステムの設計者がユーザの発話を想定して作成するものである。人手で作成することから、高々数百文程度であり、ここに多様な発話パターンが現れることは期待できない。一方、人手で作成しているため、既にクラスは付与されているものとする。

(3) 類似ドメイン大規模コーパス

既存の大規模コーパスを利用したり、Web などから収集したりして用意する。数万から数百万文程度の大規模なコーパスであり、多様な発話パターンが出現すると期待できる。一方、ドメインが検索対象データベースとは必ずしも一致しないため、データベースに登録されているドメイン固有語が全て出現するとは限らない。

クラス N-gram 言語モデルのクラスとそのクラスに属するドメイン固有語は、検索対象データベースの属性に基づき決定する。したがって、ドメイン固有語辞書はクラスごとに作成する。

3.2 適切なクラス付与

コーパスへのクラス付与は、単語の出現確率に大きく影響するため、適切に行わなければならない。クラス N-gram モデルにおける単語 w_i の出現確率 $P(w_i)$ は、クラス内確率 $P(w_i|c_i)$ とそのクラスの連鎖確率 $P(c_i|c_{i-1}c_{i-2}\dots c_{i-N+1})$ の積として、次式のように表される。

$$P(w_i) = P(w_i|c_i)P(c_i|c_{i-1}c_{i-2}\dots c_{i-N+1})$$

ただし、 i は単語列における位置、 c_i は単語 w_i が属するクラス、 N は N-gram モデルの N である。本研究では、ドメイン固有語をクラス内単語とし、その他の単語を 1 クラス 1 単語としている。コーパスへの適切なクラス付与とは、以下の 2 点を満たす必要がある。

(1) ドメイン固有語が出現する文脈として適切な箇所にクラスが付与されている

ドメイン固有語が出現する文脈として不適切な箇所にクラスが付与されていると、その不適切な文脈でのクラスの連鎖確率が高くなる。これにより音声認識誤りが起こりやすくなる。

(2) 適切な文脈には可能な限りクラスが付与されている

適切な文脈にクラスが付与されていたとしても、その数が少なければ、クラスの連鎖確率は低くなる。これ

により、ドメイン固有語が認識されにくくなる。

単純なクラス付与手法として、検索対象データベース中の単語と、類似ドメイン大規模コーパス中の単語の一致をとって、クラスを付与する方法があるが、これでは上記 2 点を満たせない。単純なクラス付与の例を図 1 に示す。図 1 は、レストラン検索のデータベースに、「そば」が食べ物を表す FOOD クラスとして登録されている場合である。「FOOD」を付与する際には、図 1 の (1) のように FOOD クラスが現れる文脈として適切な箇所に付与するのが望ましい。しかし、データベースとコーパスの一致をとってクラスを付与する場合、データベースの単語がコーパス中で意図した意味で出現するとは限らない。図 1 の (2) は、「近く」を意味する「そば」に「FOOD」が付与してしまった例を示している。これは、FOOD クラスのドメイン固有語が出現する文脈として適切ではないため、音声認識誤りの原因となりうる。他方、データベースに出現しないが、適切な箇所に出現する単語にクラスが付与できないといった問題もある。図 1 の (3) の「レシピ」がそれにあたる。「レシピ」は食べ物ではないため、データベースには登録されていないが、図 1 の (3) に示すような文脈においては、FOOD を付与することが望ましい。例えば「... おすすめのケーキがあれば教えて...」のように、「レシピ」を FOOD クラスの「ケーキ」に置き換えても文脈として適切であるからである。単純なクラス付与では、図 1 の (3) のようなことは頻繁に起こりうる。

本研究は、ドメイン固有語が出現する文脈に基づいてクラスを付与することで、コーパスに適切なクラス付与を行い、ドメイン固有語が正しく認識できるクラス N-gram モデルの構築を目指す。

4. コーパスへのクラス付与に基づくクラス N-gram モデルの構築

提案する機械学習を用いたクラス付与では、コーパスと検索対象データベースを入力とし、これらからクラス付き大規模コーパスを作成する。クラス N-gram モデル構築の概要を図 2 に示す。クラスが付与された大規模コーパスと、当該ドメインのクラス付き小規模コーパスからクラス N-gram モデルを構築する。

機械学習を用いて、ドメイン固有語が出現する文脈に基づきクラスを付与し、そのクラス付きコーパスからクラス N-gram モデルを構築する。提案するクラス N-gram モデル構築の手順を示す。

(1) 検索対象データベースの属性からクラスを設定する。各クラスに属する単語を記したドメイン固有語辞書を作成する。

(2) コーパス (類似ドメイン大規模コーパスまたは当該ドメイン小規模コーパス) 中のドメイン固有語辞書に出

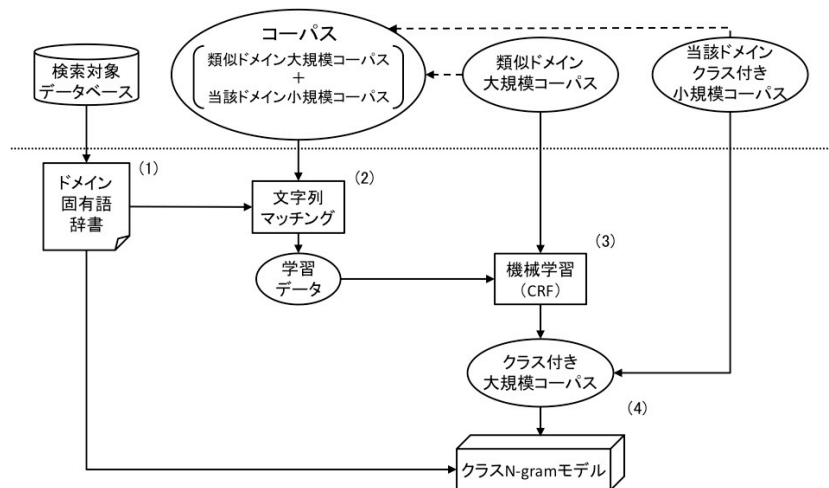


図2 クラス N-gram モデル構築の概要

現する単語に対してクラスを付与し、学習データを作成する。

- (3) 学習データの作成に使用した類似ドメイン大規模コーパスを入力とし、これに対して機械学習モデルを用いてクラスを付与する。
- (4) 作成したクラス付き大規模コーパスと当該ドメインのクラス付き小規模コーパスを混合し、クラス N-gram モデルを構築する。

3.2節で述べたように、クラス N-gram モデル構築のためのクラス付きコーパスは、クラスが出現する文脈の単語に、可能な限りクラスが付与されていることが望ましい。そこで、まず、コーパスと辞書の一致をとってクラスを付与したのからドメイン固有語が出現する文脈を学習する。学習結果に基づき、(3)で改めて類似ドメインのコーパスにクラスを付与する。文脈に基づくクラス付与により、類似ドメインのコーパスに対して、ドメイン固有語辞書に存在しない単語にもクラスを付与し、適切な文脈へのクラス付与数の増加を図る。

4.1 学習データの作成

本章では、クラス付与に用いる機械学習のための学習データの作成法について述べる。音声対話システム構築時には、学習に利用できるクラス付き大規模コーパスは一般には存在しない。そこで、3.1節で述べた、構築時に利用できる3つの資源から学習データを作成する。本研究では学習データを、コーパスとドメイン固有語辞書に出現する単語にクラス付与を行い作成する。後述する実験では、当該ドメイン小規模コーパスと類似ドメイン大規模コーパスからそれぞれ学習データを作成する。

4.2 機械学習を用いたコーパスへのクラス付与

本研究では、機械学習を用いて、ドメイン固有語が出現

する文脈にクラスを付与する。機械学習を用いてドメイン固有語が出現する文脈を学習し、それに基づきクラスを付与することで、ドメイン固有語が出現する文脈として適切な箇所にクラスを付与する。

機械学習モデルには Conditional Random Fields (CRF) [9] を用いる。ドメイン固有語の出現する文脈に基づいてクラスを付与する場合は、前後の単語のクラスも考慮する必要がある。これは系列ラベリング問題であるため、それに適した機械学習手法である CRF を用いる。CRF は、系列全体としての確からしさを学習するので、これによって文脈を扱える。コーパスに付与するクラスを表すラベルは、IOB2を用いる。つまり、ドメイン固有語を表すクラスにそれぞれ 'B-<CLASS>', 'I-<CLASS>' ラベルが存在する。<CLASS> には設定した各クラスが入る。ドメイン固有語以外の単語には 'O' ラベルを付与する。

使用する特徴量は、ドメイン固有語が出現する文脈を学習するため、クラス付与対象単語の前後の単語と品詞を用いる。同一クラスに属するドメイン固有語は、似た文脈で出現すると考えられる。そこで、対象単語の前後の単語と品詞が文脈を表すと考え、これを特徴量として用いる。つまり、単語 w_i にクラスを付与する場合、 w_i 周辺の単語と品詞で判断する。

単語 w_i にクラス付与する際に使用する特徴量として、その前後 *window* の範囲の単語 1-gram、品詞 1-gram と w_i の品詞を用いる。後述する実験では、*window* を変化させて音声認識率が最も良くなる *window* の値を調査する。

4.3 クラス N-gram モデルの構築

クラスを付与した類似ドメイン大規模コーパスとクラス付き当該ドメイン小規模コーパスを混ぜ合わせ、クラス N-gram モデルを構築する。当該ドメイン小規模コーパスと類似ドメイン大規模コーパスをそのまま混合すると、当

表1 設定したクラス

クラス	単語の例	登録単語数
FOOD	ラーメン, ホルモン	83
GENRE	中華, 和食	23
LOCATION	栄, 北区	20
STATION	栄, 名古屋	231

該ドメイン小規模コーパスの影響がほとんど現れない。そこで、当該ドメイン小規模コーパス中の各文を複製して、類似ドメイン大規模コーパスと同程度のサイズにしてから混合し、クラス N-gram モデルを構築した。クラス内確率は、検索対象データベースにおける各ドメイン固有語の頻度をもとに設定した。データベースでのドメイン固有語の出現頻度は、ユーザ発話でのドメイン固有語の出現頻度がある程度反映されていると考えられる。つまり、データベースに頻出するドメイン固有語は、ユーザ発話においても頻出すると考えられる。そこで、クラス c_i のクラス内単語 dsw_j のデータベースでの出現回数が $f(dsw_j)$ である場合、 dsw_j のクラス内確率を以下のように設定する。

$$P(dsw_j|c_i) = \frac{f(dsw_j)}{\sum_{k=1}^{N_i} f(dsw_k)}$$

ただし、 N_i はクラス c_i の語彙サイズを表す。

5. 音声認識実験

本章では、以下を変えた際の音声認識の比較を行う。

- (1) クラス付与手法
- (2) CRF の学習データ中のクラスが含まれる文の割合
- (3) CRF の特徴量の *window*

本章の実験に共通する実験条件について述べる。ドメインは、愛知県のレストラン検索である。このドメインの検索対象データベースから、ドメイン固有語の属する4クラスを設定した。設定したクラスと属するドメイン固有語の例を表1に示す。類似ドメイン大規模コーパスとして、Yahoo!知恵袋の中カテゴリ「料理, グルメ, レシピ」に属する1,208,269文を使用した。当該ドメイン小規模コーパスには、愛知県のレストラン検索での発話を想定することを教示し、10名から1名につき平均13文を収集した合計132文を用いた。このコーパスには、あらかじめドメイン固有語辞書と一致する単語にクラスを付与した。類似ドメイン大規模コーパスと混ぜ合わせる際は、各文を1万回複製し混ぜ合わせた。評価データには、言語モデルの構築とは別に、文献[10]のシステムを用いて収集したレストラン検索ドメインの発話を使用した。収集した対話は、全部で120対話(被験者30名, 各4対話)である。今回作成した言語モデルは、「マクドナルド」を「マック」「マクド」と呼ぶような省略語を考慮していない。そのため、ドメイン固有語へのクラス付与を行い、言語モデルを作成したことによる効果のみを確認するため、収集した発話から雑音や店名に関する発話を除いた4,480発話(14,554単語, うちド

表2 クラス付与手法の違いによる音声認識率の比較

モデル名	類似ドメイン大規模コーパスへのクラス付与数	WER [%]	
		全単語	ドメイン固有語
BaseLine	165,160	35.46	24.6
文献 [2] の手法	1,494,481	38.34	23.6
提案手法	50,116	35.46	23.3

- (1) 他に金沢で美味しい和菓子/B-FOOD屋ありますか。
- (2) 食べ物/B-FOOD屋といっても詳しく言うと某チェーン店のです。

図3 提案手法によってドメイン固有語辞書にない単語にクラスが付与できた例

メイン固有語1,454)を使用した。評価指標には評価データ中の全単語とドメイン固有語それぞれに対する Word Error Rate (WER) を用いた。形態素解析には Mecab^{*1}, CRF には CRF++^{*2}, 音声認識には Julius^{*3}を、それぞれ使用した。

5.1 機械学習を用いたクラス付与の有効性の確認

5.1.1 実験条件

コーパス中の適切な文脈へのクラス付与の重要性を示すため、クラス付与手法の異なる3種類のクラス N-gram モデルの音声認識率を比較した。

(1) BaseLine

類似ドメイン大規模コーパスと、検索対象データベースから作成したドメイン固有語辞書の文字列マッチングにより、クラスを付与する。

(2) 文献 [2] の手法

まず、類似ドメイン大規模コーパス中の適切な文脈に出現する単語を機械学習により抽出する [2]。それとドメイン固有語辞書の単語を用いて、類似ドメイン大規模コーパスとの文字列マッチングによりクラスを付与する。ここで用いる機械学習の学習データには、類似ドメイン大規模コーパスとドメイン固有語辞書の一致をとってクラス付与したコーパスを用いる。

(3) 提案手法

提案するクラス付与手法により類似ドメイン大規模コーパスにクラスを付与する。機械学習の学習データには、類似ドメイン大規模コーパスとドメイン固有語辞書の一致をとってクラス付与したコーパスを用いる。CRF の特徴量には、クラス付与対象単語 w_i の前後3単語 1-gram と前後3品詞 1-gram, そして w_i の品詞を用いた。

5.1.2 実験結果

類似ドメイン大規模コーパスへのクラス付与手法を変えた際の音声認識率を表2に示す。また、提案手法のときのクラスが付与された文の例を図3に示す。表2の類似ドメ

*1 <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

*2 <http://crfpp.googlecode.com/svn/trunk/doc/index.html>

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

表3 rate による音声認識率の変化

学習データ	パターン	rate (学習データの合計文数)	類似ドメイン大規模コーパスへのクラス付与数	WER [%]	
				全単語	ドメイン固有語
当該ドメイン 小規模コーパス	1	default (132)	58,791	35.30	22.4
	1	0.5 (76)	144,843	35.21	21.9
		1.0 (38)	569,042	34.91	21.4
	2	0.9982 (516,339)	483,965	34.85	21.0
類似ドメイン 大規模コーパス	1	default (1,208,269)	50,116	35.46	23.3
	1	0.5 (265,860)	146,145	35.56	22.8
	1	1.0 (132,930)	341,492	35.51	21.7

イン大規模コーパスへのクラス付与数を見ると、文献 [2] の手法が他の 2 つに比べ、非常に多いことがわかる。これは、「辛いのが好きなので」という文脈の「の」を FOOD クラスのドメイン固有語として抽出し、これを用いてクラス付与を行っていたためである。「辛いのが...」の「の」は、文脈的に FOOD クラスのドメイン固有語として適切であるが、他の箇所でも出現する多くの「の」は FOOD クラスとして不適切である。これらにクラスを付与したため、文献 [2] の手法においては、全単語の WER が他の 2 つのモデルに比べ高くなった。これから、クラス付与箇所が多すぎて適切な箇所に付与しなければ、音声認識率の向上には繋がらないと確認できる。次に、BaseLine と提案手法を比較すると、提案手法の方がクラス付与箇所は少ないにも関わらず、ドメイン固有語の WER が低い。これは図 3 に示すように、ドメイン固有語辞書には存在しないが、ドメイン固有語が出現する文脈として適切な箇所の単語にクラスが付与できたためと考えられる。全体としてみたクラス付与箇所は少ないが、適切な箇所に付与できたため、ドメイン固有語の WER が低下したと考えられる。これから、適切な箇所へのクラス付与が重要であり、提案手法によってこれが実現できていることがわかる。

5.2 CRF の学習データと音声認識率の関係

5.2.1 実験条件

適切な箇所へのクラス付与箇所が多いほど、ドメイン固有語の音声認識率が向上することを示す。さらに、学習データによってドメイン固有語の WER に影響が出るか調べるため、学習データとして類似ドメイン大規模コーパスと当該ドメイン小規模コーパスをそれぞれ使用した。

ドメイン固有語が出現する文脈として適切な箇所にクラスが付与できるならば、コーパスへのクラス付与箇所が多いほどドメイン固有語は認識されやすくなる。そこで、学習データ中のクラスが付与された単語を含む文(クラス文)の割合 (rate) を増加させることで、機械学習にクラスの出現確率が高くなるように学習させ、クラス付与箇所の増加を図る。初期の学習データでは、rate は低い。具体的には、類似ドメイン大規模コーパスの場合は、

$$C_{class} : C_{other} = 132,930 : 1,075,339$$

$$(rate = 0.110)$$

当該ドメイン小規模コーパスの場合は、

$$C_{class} : C_{other} = 38 : 94$$

$$(rate = 0.3)$$

であった。ただし、 C_{class} は学習データ中のクラスクラス文の数、 C_{other} は学習データ中のクラスが付与されなかった文(非クラス文)の数である。そのため、この初期の学習データを用いて機械学習を行うと、学習コーパス全体の中で、クラスが出現する確率が低く抑えられてしまう。そこで、次の 2 パターンの方法で、学習データ中の rate を増加させた。

パターン 1: 学習データ中の非クラス文を減らす。

$$rate = \frac{C_{class}}{C_{class} + C_{other}}$$

となるように、非クラス文を減らした。クラス文については、類似ドメイン大規模コーパスに関しては 132,930 文、当該ドメイン小規模コーパスに関しては 38 文をそのまま使用した。

パターン 2: 学習データ中のクラスが文を増やす。

学習データ中のクラスが付与された単語を、ドメイン固有語辞書の単語と置き換えた文も学習データとすることで、クラス文を増やした。例として「うどん/FOOD 屋 教えて」という文が学習データにあり、FOOD クラスのドメイン固有語として「ラーメン」、「焼肉」があったとする。このとき、学習データに「ラーメン/FOOD 屋 教えて」、「焼肉/FOOD 屋 教えて」が追加される。ただし、パターン 2 は、学習データが類似ドメイン大規模コーパスの場合に行うと、学習データの量が急増するため、今回は学習データが当該ドメイン小規模コーパスの場合に對してのみ行った。特徴量には、クラス付与対象単語 w_i の前後 3 単語 1-gram と前後 3 品詞 1-gram, w_i の品詞を用いた。

5.2.2 実験結果

表 3 に学習データに用いたコーパスと、学習データの作成方法を変えた際の類似ドメイン大規模コーパスへのク

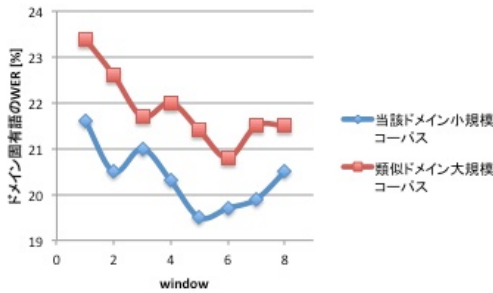


図4 window による音声認識率の変化

ラス付与数と音声認識率を示す。ただし、表3の *rate* が default のときの結果は、*rate* を増加させる処理を行っていない。特に、学習データが類似ドメイン大規模コーパスの場合の結果は、表2の提案手法と同じである。まず、全単語に対する WER を見ると、条件に関わらず、ほとんど変化がないことがわかる。次に、学習データに使用したコーパスの種類によらず、*rate* が大きくなるに従って、クラス付与数が増加することが確認できる。*rate* が大きくなると、クラスが出現する文脈を強く学習することになるので、クラス付与数が増加したと考えられる。

パターン1の場合においてドメイン固有語の WER を見ると、クラス付与数が増加するに従い、WER は低下する傾向にある。これは、クラス N-gram モデルの構築に用いるクラス付きコーパスのクラス付与数が増えることで、各クラスの出現確率が上がるため、ドメイン固有語を認識しやすくなったと考えられる。この結果から、適切な箇所にはクラスが付与できるのであれば、クラス付与箇所が多いほうがドメイン固有語を認識しやすいと言える。

最後に、学習データが当該ドメイン小規模コーパスの際の、学習データ中のクラス文の増加パターンの違い(パターン1とパターン2)によるドメイン固有語の WER を比較する。比較は、*rate* = 1.0 と *rate* = 0.9982 のときである。これらを比較すると *rate* = 0.9982 の方が *rate* = 1.0 のドメイン固有語の WER より低い。これはパターン2の増加方法の方が、よりドメイン固有語が出現する文脈を学習しやすいからであると考えられる。パターン2の学習データでは、同じ文脈でも異なるドメイン固有語が出現する。そのため、パターン1の学習よりも、適切な文脈を強く学習したためだと考えられる。その結果、クラス付与箇所が少なくても、パターン2の方がドメイン固有語の WER は低くなった。

以上の結果から、学習データ中のクラス文の割合を大きくすることで、クラス付与箇所が増加することを確認した。それに伴い、ドメイン固有語の WER が低下することを確認した。

- (1) 正解: 「名東区の居酒屋を教えてください」
(名東区: LOCATION, 居酒屋: GENRE)
- BaseLine: 「麺と具の居酒屋を教えてください」
 - win7_in-domain: 「名東区の居酒屋を教えてください」
 - win6_yahoo: 「名東区の居酒屋を教えてください」
- (2) 正解: 「栄の松屋デパートの近くのお店」
(栄: STATION)
- BaseLine: 「魚を混ぜ揚げ方の近くのお店」
 - win7_in-domain: 「栄の松屋で簡単近くのお店」
 - win6_yahoo: 「栄の松屋で簡単に近くのお店」
- (3) 正解: 「栄にある居酒屋のお店を教えてください」
(栄: STATION, 居酒屋: GENRE)
- BaseLine: 「栄にある田舎の店をおしえて」
 - win7_in-domain: 「栄にある居酒屋のお店をおしえて」
 - win6_yahoo: 「栄にある居酒屋のお店をおしえて」

図5 認識結果の例

5.3 CRF の特徴量と音声認識率の関係

5.3.1 実験条件

CRF において、クラス付与対象単語 w_i の前後何単語・品詞を特徴量として用いるか決定する *window* を変化させて実験を行った。特徴量は、クラス付与対象単語 w_i 前後の単語・品詞数 *window* を、*window* = 1 ~ 8 として実験を行った。学習データには、当該ドメイン小規模コーパスと類似ドメイン大規模コーパスを用いた。当該ドメイン小規模コーパスは5.2節で述べたパターン2の方法で *rate* を増加させ *rate* = 0.9982 とした。類似ドメイン大規模コーパスは5.2節で述べたパターン1の方法で *rate* を増加させ *rate* = 1.0 とした。

5.3.2 実験結果

図4に特徴量の *window* を変化させた時のドメイン固有語の WER を示す。横軸が *window*、縦軸がドメイン固有語の WER をそれぞれ表す。青線が、学習データに当該ドメイン小規模コーパスを使用した場合、赤線が、学習データに類似ドメイン大規模コーパスを用いた場合である。同一の学習データを用いた場合、*window* を変化させたときに認識率に統計的に有意な差があるか確認するため、2群比率の差検定を有意水準5%で行った。その結果、学習データに関わらず、同一の学習データを用いたときには、*window* を変化させても統計的に有意な差があるとは言えなかった。ドメイン固有語の WER が最小になったのは、学習データが当該ドメイン小規模コーパスの場合、*window* = 5 のときの19.5%である。学習データが類似ドメイン大規模コーパスの場合、*window* = 6 のときの20.8%である。

学習データを変えたときの、ドメイン固有語の WER の統計的な有意差は見られなかった。なお、当該ドメイン小規模コーパスを学習に用いた場合の認識性能は、その収集

方法や内容に依存すると予想される。つまり、今回の実験のように、当該ドメイン小規模コーパスを学習に用いた方が常に良くなるとは限らない。このことから、安定してドメイン固有語の認識率の向上が期待できる、類似ドメイン大規模コーパスを用いるのが良いと言える。

最後に、音声認識結果の例を図5に示す。次の3つの条件での認識結果を比較する。

- BaseLine : 5.1 節の BaseLine
- win5_in-domain : 5.3 節の、学習データが当該ドメイン小規模コーパスかつ $window = 5$ の時のモデル
- win6_yahoo : 5.3 節の、学習データが類似ドメイン大規模コーパスかつ $window = 6$ の時のモデル

図5の(1)を見ると、LOCATION クラスのドメイン固有語「名東区」に関して、BaseLineでは「麵と具」と誤認識されていることがわかる。これに対して、提案手法によって構築した win5_in-domain, win6_yahoo は正しく「名東区」を認識できていることがわかる。同様に図5の(3)を見ると、GENRE クラスのドメイン固有語「居酒屋」に関して、BaseLineでは「田舎」と誤認識されているが、win5_in-domain, win6_yahoo は正しく「居酒屋」を認識できている。これらの結果も示すように、文脈に基づきクラス付与を行い構築したクラス N-gram モデルによってドメイン固有語を認識できることがわかる。BaseLine と win5_in-domain, win6_yahoo のドメイン固有語の WER それぞれに対して、2群比率の差検定を有意水準 5%で行ったところ、統計的に有意な差があることを確認した。

6. おわりに

本稿では、音声対話システム構築時に利用できるコーパスと検索対象データベースから、CRFの学習データを作成した。それに基づきコーパスに対してクラス付与を行い、そのコーパスから構築したクラス N-gram モデルで、ドメイン固有語の音声認識率が向上することを確認した。実験により、CRFの学習データは、その作成元のコーパスに関わらず、クラス文の割合が高い場合にコーパスへのクラス付与数が増加することを確認した。ベースラインと比較し、そのクラス付きコーパスから構築したクラス N-gram モデルのドメイン固有語の音声認識率が、最大で5.1%向上することを確認した。

謝辞 言語モデルの作成にはヤフー株式会社が国立情報学研究所に提供した Yahoo!知恵袋データを利用した。本研究の一部は、JST 戦略的創造研究推進事業さきがけの支援を受けた。

参考文献

- [1] 駒谷和範, 河原達也, 清田陽司, 黒橋禎夫, FUNG Pascale. 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, 音声言語情報処理, Vol. 2001, No. 123, pp. 177-182, 2001-12-20.
- [2] 森祥二郎, 駒谷和範, 佐藤理史. 大規模コーパスへのクラス付与に基づく音声対話システム用言語モデルの構築. 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2012, No. 16, pp. 1-8, 2012-07-12.
- [3] 森信介, 西村雅史, 伊東伸泰. クラスに基づく言語モデルのための単語クラスターリング. 情報処理学会論文誌, Vol. 38, No. 11, pp. 2200-2208, 1997-11-15.
- [4] 池谷晴生, 福田隆, 山田博文, 桂田浩一, 新田恒雄. 意味属性を利用したクラス n-gram 言語モデルの評価. 電子情報通信学会技術研究報告. SP, 音声, Vol. 104, No. 542, pp. 31-36, 2004-12-14.
- [5] 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. D, 情報・システム, Vol. 90, No. 11, pp. 3024-3032, 2007-11-01.
- [6] Abhinav Sethy, Panayiotis G. Georgiou, Bhuvana Ramabhadran, and Shrikanth Narayanan. An iterative relative entropy minimization-based data selection approach for n-gram model adaptation. *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, Vol. 17, No. 1, pp. 13-23, Jan 2009.
- [7] 内元清貴, 馬青, 村田真樹, 小作浩美, 内山将夫, 井佐原均. 最大エントロピーモデルと書き換え規則に基づく固有表現抽出. 自然言語処理, Vol. 7, No. 2, pp. 63-90, 2000-04-10.
- [8] 山田寛康, 工藤拓, 松本裕治. Support vector machine を用いた日本語固有表現抽出. 情報処理学会論文誌, Vol. 43, No. 1, pp. 44-53, 2002-01-15.
- [9] John Lafferty, Andrew McCallum, and Fernando C.N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *International Conference on Machine Learning*, Vol. 18, pp. 282-289, 2001.
- [10] 西村良太, 駒谷和範. データベース検索音声対話システムにおける対話状態の推定. 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2012, No. 20, pp. 1-7, 2012-01-27.