

ドメイン固有語の認識率向上を目指した音声対話システム用言語モデルの構築

森 祥二郎[†]駒谷 和範[†]佐藤 理史[†][†]名古屋大学大学院 工学研究科 電子情報システム専攻

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、タスク遂行のために、ドメイン固有語の認識が必須である。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。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。新たな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は限らない。そのため、そのシステムの言語モデルは、当該ドメインの大規模コーパスを用いずに構築する必要がある。

本研究では、当該ドメインの小規模コーパス（数百文程度）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。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。ドメイン固有語をクラスとすることで、コーパス中の適切な箇所に入力できるようにする。これにより、コーパス中に直接出現しなくても、ドメイン固有語を認識可能な言語モデルを構築する。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。

このクラス付与手法を工夫することで、ドメイン固有語の音声認識率を向上させる。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として適切な箇所を学習する。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手法を採った [2]。これに対して本稿では、単語自体ではなく、ドメイン固有語が出現する文脈を重視したクラス付与を行う。

2. 単語に基づくクラス付与の問題点

コーパスにクラスを付与する際には、以下の2点を満たす必要がある。

要求仕様 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、その文脈でのクラスの連鎖確率が不当に高くなる。これにより音声認識誤りが起こりやすくなる。

要求仕様 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。

不適切なクラス付与の例を図1に示す。図1は、レストラン検索のドメイン固有語辞書に、「そば」が食べ物

- (1) ... のおすすめのそば/FOOD を教えて ...
- (2) ... 駅のそば/FOOD にある ...
- (3) ... おいしい レシピ を教えて ...

図1: 単語の一致に基づくクラス付与の例

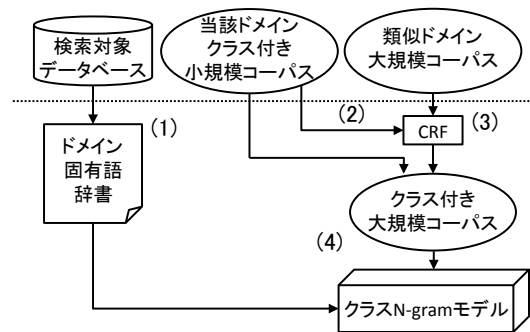


図2: クラス N-gram モデル構築の概要

を表す FOOD クラスとして登録されている場合である。類似ドメイン大規模コーパス中の単語と、ドメイン固有語辞書中の単語「そば」の一致をとってクラスを付与すると、図1の(1)のような適切な箇所だけでなく、図1の(2)のような不適切な箇所にもクラスが付与される。これは要求仕様1を満たさない。特に文献 [2] の手法では、ドメイン固有語として不適切な単語を抽出してしまうと、図1の(2)のような不適切な箇所に、大量にクラスを付与する問題があった。一方、図1の(3)の「レシピ」のような、ドメイン固有語辞書に出現しないが、適切な箇所に出現する単語にクラスが付与できると、要求仕様2が満たされる。このような単語には、単語の一致をとる手法ではクラスを付与できない。特に、ドメイン固有語辞書をデータベースのみから作成した場合には、このようなことは頻繁に起こりうる。

3. 文脈に基づくコーパスへのクラス付与

提案するクラス付与手法の概要を図2に示す。

- (1) 検索対象データベースの属性に基づきクラスを設定する。各クラスに属する単語を含むドメイン固有語辞書を作成する。
- (2) 当該ドメインクラス付き小規模コーパスを学習データとして Conditional Random Fields (CRF) を学習する。
- (3) CRF を用いて類似ドメイン大規模コーパスにクラスを付与する。
- (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。

Constructing Language Model for Spoken Dialogue Systems to Improve ASR Accuracy of Domain-Specific Words. Shojiro Mori, Kazunori Komatani, and Satoshi Sato (Nagoya Univ.)

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	北区, 新栄	20
STATION	名古屋, 伏見	231

表 2: クラス付与手法の違いによる WER の比較 [%]

モデル	全単語	DSW
文字列マッチング	35.46	24.6
DSW に基づく手法	38.34	23.6
提案手法	34.67	19.5

DSW : ドメイン固有語

要求仕様 1 を満たすためには、前後の単語のクラスを考慮する必要がある。そこで、CRF を用いて、クラス系列全体としての確からしさを学習し、文脈を扱う。コーパスに付与するクラスを表すラベルは、IOB2 を用いる。

要求仕様 2 を満たすため、CRF の学習時には、学習データ中のクラスが付与された単語を含む文を増加させる。具体的には、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を、学習データに加えることで、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」を追加する。

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。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、予備実験により $window = 5$ とした。

4. 評価実験

提案手法と次の 2 種類のベースライン手法でクラスを付与し、クラス N-gram モデルの音声認識率を比較した。

1. **文字列マッチング** : 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との、文字列マッチングによりクラスを付与する。
2. **ドメイン固有語に基づく手法 [2]** : 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。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。

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、10 名から 1 名につき平均 13 文ずつ収集した、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。クラス内確率は、検索対象

- (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)

- 文字列マッチング : 「麺 と 具 の 居酒屋 を 教えてください」
- 提案手法 : 「名東区 の 居酒屋 を 教えてください」

- (2) 正解 : 「栄 に ある 居酒屋 の お店 を 教えて」
(栄 : STATION, 居酒屋 : GENRE)

- 文字列マッチング : 「栄 に ある 田舎 の 店 を おしえて」
- 提案手法 : 「栄 に ある 居酒屋 の お店 を おしえて」

図 3: 認識結果の例

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。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を $f(dsw_j)$ とし、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。

評価データには、言語モデルの構築とは別に収集した、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。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, CRF には CRF++, 音声認識には Julius を、それぞれ使用した。

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。なお、提案手法により、図 1 の (3) の「レシピ」のような箇所にもクラスが付与されたことを確認した。まず、全単語の WER を見ると、ドメイン固有語に基づく手法で値が大きくなっている。これは、不適切な箇所にクラスが大量に付与されたためである。次に、ドメイン固有語の WER を見ると、提案手法は、他の 2 つに比べ低い。実際、図 3 に示す認識結果の例を見ると、(1) の例のように、LOCATION クラスのドメイン固有語「名東区」が、文字列マッチングでは「麺 と 具」と誤って認識されている。同様に、(2) から、GENRE クラスのドメイン固有語「居酒屋」も、提案手法では、正しく認識されている。以上から、コーパス中のドメイン固有語が出現する適切な文脈にクラスを付与し、そのクラス付きコーパスから構築したクラス N-gram モデルにより、ドメイン固有語の認識率が向上することがわかる。

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。

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, 2001-SLP-39, pp. 177-182, 2001.
- [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, Vol. 2012-SLP-92, No. 16, 2012.