Vol.2012-SLP-92 No.16
2012/7/21

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

森 祥二郎[1]　駒谷 和範[1]　佐藤 理史[1]

概要：音声対話システムでは地名などの固有名詞（内容語）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 モデルの自動作成を行う．これにはクラスが多数付与された大規模コーパスが必要であるが，個別の音声対話システムのドメインにおいて大規模コーパスの存在を仮定するのは現実的ではない．そこで我々は，類似ドメインの大規模コーパスを用い，その中で，検索対象データベース中の内容語と一致する部分を種とした機械学習を行うことで，徐々にクラス付与箇所を増加させるというアプローチを採る．これにより，内容語周辺の多様な発話パターンを認識可能な言語モデルの構築を目指す．評価実験により，提案する枠組みにより，内容語の認識率が向上する可能性を示す．

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

Constructing Language Model for Spoken Dialogue Systems based on Assigning Semantic Classes to Large-Scale Corpus

SHOJIRO MORI[1]　KAZUNORI KOMATANI[1]　SATOSHI SATO[1]

Abstract: Content words such as proper nouns must be correctly recognized in spoken dialogue systems. We are trying to automatically construct a class N-gram model to recognize user utterances containing such content words. Although a large-scale corpus with the classes is required to construct the model, it is not realistic to assume that such a corpus is available for each individual domain of the target spoken dialogue system. We then use a similar-domain corpus and assign semantic classes to it via machine learning in a bootstrapping manner. The experimental evaluation showed that our proposed framework can improve ASR accuracy of content 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，内容語の認識が重要である．ここでの内容語とは，地名等のそのドメインに固有な名詞を指す．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) 当該ドメインの内容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．内容語に関しては，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ことを仮定するのは現実的でない．

単純に，当該ドメインの音声認識用の言語モデルを作成する方法として，内容語を文法カテゴリとした文法ルール

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

を人手で記述することが挙げられる．これには，ユーザ発話の多様な発話パターンの認識が困難という問題がある．

一方で，多様な発話パターンの認識には，統計的言語モデル (N-gram モデル) により，高い音声認識率を得られることが知られている．統計的言語モデルの中でも，クラス N-gram モデルは，内容語をクラスとすることで，内容語と多様な発話パターンの両方を認識できることが期待できる．クラス N-gram モデルの構築には，クラスが付与された大量の発話パターンを含むコーパスが必要である．学習データとしてこれを用意できれば，内容語の前後に接続する多様な発話パターンに対応したクラス N-gram モデルが構築できる．しかし，ドメインごとに，クラスが付与された大量のコーパスが存在すると仮定するのは，現実的ではない．

本研究では，当該ドメインの小規模コーパス (または当該ドメインの記述文法) 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，数百文程度なら人手で用意でき，記述文法も，単純なものであれば，用意可能であるとする．また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．これらを混合したコーパスを，当該ドメインの大規模コーパスの代わりに，クラス N-gram モデルの学習データとして用いる．文献 [1] では，当該ドメインの小規模コーパスに出現する内容語にクラスを付与し，そこに検索対象データベース中の内容語を代入している．このクラス付きコーパスと，類似ドメインの大規模コーパスとを混合し，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，類似ドメインの大規模コーパスにもクラスを付与することで，クラス (内容語) 前後の多様な発話パターンをより多く学習する．これにより，内容語周辺の多様な発話パターンの音声認識性能の向上を狙う．

類似ドメインの大規模コーパスへのクラス付与は，自動で行う必要がある．単純な自動クラス付与手法として，検索対象データベース中の内容語と完全に一致する単語へクラスを付与する方法が考えられる．この方法では，内容語がドメイン固有の単語であるため，類似ドメインのコーパス中に多くは現れず，クラスを付与できる箇所は少ない．この結果，内容語を含む多様な発話パターンは学習できない．

そこで本研究では，未知の内容語をコーパス中から抽出し，これに対してもクラスを付与する．ここで「未知」とは，検索対象データベースには存在しないものを指す．データベース中の内容語に加え，未知の内容語にもクラスを付与することで，内容語を含む発話パターンの種類を増加させる．この結果，大規模コーパスに対してクラスをより多く付与でき，このコーパスから言語モデルを構築することで，内容語を含む発話パターンの音声認識率の向上を図る．

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 章で提案手法の枠組みを述べる．提案手法の重要な要素として，3 章では大規模コーパス中の未知の内容語に，機械学習によりクラスを付与する手法について述べ，4 章でその評価実験を行う．さらに，5 章では得られてクラス付与済みコーパスを用いて音声認識実験を行い，提案する枠組みの性能やその上限について議論する．6 章では本稿のまとめと，今後の課題についてまとめる．

2. 未知の内容語へのクラス付与

2.1 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内容語が登録されている．

(2) **当該ドメインの小規模コーパス**
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) **類似ドメインの大規模コーパス**
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ことが期待できる．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内容語はほとんど出現しない．

クラス N-gram 言語モデルのクラスとそのクラスに属する内容語は，検索対象データベースの属性に基づき決定する．つまり，内容語辞書はクラスごとに作成する．

2.2 関連研究

正解クラスラベル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，未知の内容語 (固有表現) を抽出する手法が提案されている [2,3]．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から内容語を抽出する．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

パスを用意する必要がある．実際，文献 [2] では約 12,000 文，文献 [3] では約 11,000 文のラベル付きコーパスを学習に用いている．これに対して，2.1 節で述べた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が存在することを仮定するのは適当ではない．

そこで本研究では，大量の学習データを必要としない，機械学習を用いた Bootstrap 式の未知の内容語抽出を行う．Bootstrap 式に少しずつ学習を行うことで，クラスが付与される箇所を漸次的に増加させることを狙う．つまり，クラス付与対象のコーパスから，未知の内容語を抽出しながらクラスを自動付与することを目指す．

2.3 提案手法の概要

提案する Bootstrap 式クラス付与では，類似ドメインの大規模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．提案する Bootstrap 式のクラス付与の概要を図 1 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．

Bootstrap 式クラス付与では，下記の (2) 〜 (6) を繰り返すことで，対象コーパスにクラスを付与することを想定している．

(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内容語辞書を作成する．
(2) 内容語辞書に出現し，類似ドメインの大規模コーパスにも出現する単語に対して，クラスを付与しクラス付き大規模コーパスを作成する．
(3) クラス付き大規模コーパスを学習データとし，内容語と非内容語に分類する機械学習を行う．この結果，コーパスから内容語を抽出する．
(4) 抽出した未知の内容語のうち，内容語として不適当な単語 (誤抽出単語) を除去する．
(5) (4) で誤抽出単語を除去した，残りの単語を内容語辞書に追加する．
(6) (2) に戻る．

このサイクルを繰り返すことで，内容語辞書を拡張する．この結果，当初の内容語辞書には未知であった内容語にもクラスを付与したコーパスを作成する．

3. 機械学習を用いた内容語抽出

提案手法では，抽出した未知の内容語をクラス付与に用いる．クラス付与は，単純にコーパス内で内容語と一致する単語をクラスとしている．その後，新たにクラスが付与されたこのコーパスを学習データとして，新たな未知の内容語を抽出するための機械学習を行うことを想定している．そのため，内容語の抽出には以下の 2 つが重要である．

(1) **未知の内容語の抽出** (2.3 **節の** (3))
既知の内容語を利用しても，内容語が出現する新たな箇所へのクラス付与はできない．そのため，未知の内容語を抽出する必要がある．

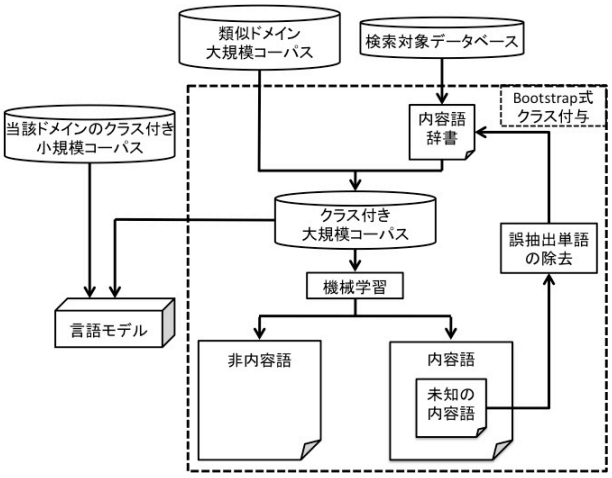

図 1 Bootstrap 式クラス付与の概念図

(2) **誤抽出単語の除去** (2.3 **節の** (4))
誤抽出された内容語を利用すると，誤った箇所にクラスを付与してしまう．このクラスを正解ラベルとした機械学習により内容語抽出を行うと，内容語の誤抽出が増加する．

本稿では，誤抽出単語の除去は未実装であるため，人手で行っている．本章では，2.3 節の (3) で述べた内容語抽出について詳しく説明する．

3.1 用いる機械学習モデル

本研究では，コーパス中の各単語を内容語とそれ以外の単語に分類し，内容語に分類された単語を抽出するという手順を内容語抽出と呼ぶ．分類には機械学習を用いる．

分類に用いる機械学習モデルとして Maximum Entropy (ME) model [4] を用いる．コーパスへのクラス付与を行うために用いられる他の機械学習モデルには，Conditional Random Fields (CRF) [5] がある．ME model を用いた理由は，CRF と比較し少量の学習データからでも，比較的高い性能が期待できるためである．

3.2 使用する特徴量

ME model に与える特徴量として，分類対象単語の前後の単語を用いる．同一クラスに属する内容語は，似たコンテキストで出現すると考えられる．そこで，対象単語の前後の単語がコンテキストを表すと考え，これを特徴量として用いる．つまり，単語 w_i が内容語かそれ以外の単語であるかは，w_i 周辺の単語で判断する．そこで，単語 w_i を分類する際に使用する特徴量として，その前後 j 単語の範囲の N-gram と w_i の品詞を用いる．本稿では $j = 3$ とした．

本稿では，前後 3 単語の範囲の 1-gram，2-gram，3-gram と w_i の品詞を特徴量とした．特徴量として用いる N-gram は，予備実験を行い決定した．具体的には，4 パターン

表 1 「吉祥寺の有名なヨウカン屋さんの名前を知りたい。」の「ヨウカン」を w_i としたときの特徴量の例

特徴量	詳細	w_i の特徴量
POS	w_i の品詞	名詞_POS
PRE	w_i の前方 1-gram	の_PRE3, 有名_PRE2, な_PRE1
FOL	w_i の後方 1-gram	屋_FOL1, さん_FOL2, の_FOL3
PREBi	w_i の前方 2-gram	の有名_PREBi2, 有名な_PREBi1
FOLBi	w_i の後方 2-gram	屋さん_FOLBi1, さんの_FOLBi2
PRETri	w_i の前方 3-gram	の有名な_PRETri1
FOLTri	w_i の後方 3-gram	屋さんの_FOLTri1

表 2 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

(1-gram のみ，2-gram のみ，3-gram のみ，1-gram・2-gram・3-gram 全て) で予備実験を行った．1-gram・2-gram・3-gram 全てを使用した場合に，抽出した未知の内容語数に対して，誤抽出単語数の率が最小となったので，これら全ての N-gram を特徴量として用いた．誤抽出単語の判断は人手で行っている．特徴量の例を表 1 に示す．

4. ME model による内容語抽出実験

3 章で述べた内容語抽出により，未知の内容語を抽出できているか確認する．

4.1 実験条件

コーパス中の単語を，内容語と非内容語に分類し，内容語として分類された単語に未知の内容語が含まれるかを確認する．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，内容語の属する 4 クラス (内容語クラス) を設定した．ME model による分類では，この 4 クラスと，内容語以外の単語を表す null クラスの合計 5 クラスに分類した．設定したクラスと属する内容語の例を表 2 に示す．登録単語数は，検索対象データベースに登録されていた単語数である．この登録単語数が，初期の内容語辞書の語彙サイズと一致する．ME model の学習データに用いるコーパスには，Yahoo!知恵袋の中カテゴリ「料理，グルメ，レシピ」(類似ドメイン大規模コーパス) の 1 万文を使用した．類似ドメイン大規模コーパスへのクラス付与は，データベース中の内容語と完全に一致する単語へクラスを付与する方法を採った．この結果，クラスが付与された単語は内容語クラスの単語，クラスが付与されなかった単語は null クラスの単語とする．

機械学習に用いる学習データを作成する際に，調整パラメータ n を導入した．初期の学習データには，内容語クラスの単語が少ない．具体的には，

$$C_{null} : C_{content} = 168,688 : 699$$

であった．ただし，C_{null} はクラス付きコーパス中の null クラスの単語数，$C_{content}$ はクラス付きコーパス中の内容語クラスの単語数である．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，

$$C_{null} : C_{content} = n : 1$$

となるように，内容語クラスの単語全体を，整数回複製し，その数を増加させた．null クラスの単語に関しては，コーパス中に出現するものをそのまま使用した．

形態素解析には Mecab[*1] を使用した．ME model の作成には mallet[*2] を使用した．学習した ME model を用いて，学習に使用したものと同じ Yahoo!知恵袋の中カテゴリ「料理，グルメ，レシピ」の 1 万文に対して内容語抽出を行った．

4.2 実験結果

まず，n を変化させることで，内容語抽出結果を操作できることを示す．表 3 に，内容語抽出結果を示す．抽出単語数は，提案手法によって抽出した単語の異なり数を示す．既知の内容語数は，データベース中の内容語が出現した数を表す．正解単語数は，未知の内容語を人手で確認し，内容語として妥当であると判断した単語数を表す．誤抽出単語数は，提案手法により未知の内容語として抽出されたが，人手で確認した結果，内容語として不適当であると判断した単語数を表す．表 3 を見ると，n が小さいほど，正解単語数，誤抽出単語数ともに多く，n が大きいほど，正解単語数，誤抽出単語数ともに少なくなることがわかる．これは，n が増大するに従い，学習データに出現する内容語クラスの単語の数が減少することで，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からである．

正解単語数は多いほど良く，誤抽出単語数は少ないほど良いが，調整パラメータ n では，両方を同時に実現することはできていない．そのため，正解単語数を増やしつつ，誤抽出単語数を減らすことは，n を変化させるだけでは実現できない．したがって，抽出する正解単語数を増やす，または，誤抽出単語数を減らすには，n の変更とは別の方法を取る必要がある．

次に，実際に抽出された正解単語を示し，内容語として妥当である単語が抽出できたことを示す．表 4 に，$n = 1$ のときの正解単語と誤抽出単語の一部を示す．表 4 を見ると，FOOD クラスとして抽出された「モッツァレラチーズ」のように，内容語として妥当な単語が抽出できていることがわかる．

さらに，表 4 中で示されている例に関して，誤抽出単語とその単語が抽出されたコンテキストを示すことで，現在

[*1] http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html
[*2] http://mallet.cs.umass.edu/

表 3 内容語抽出結果

n	抽出単語数	既知の内容語数	未知の内容語数	
			正解単語数	誤抽出単語数
1	343	55	178	110
10	220	55	108	57
60	115	51	43	21
120	82	50	18	14
241	63	50	7	6

表 4 抽出した未知の内容語の例

クラス	正解単語	誤抽出単語
FOOD	炒飯，コールスローサラダ 鍋，モッツァレラチーズ	あれ，全て，バリラ 焼き，フランス
GENRE	ティラミス，スパゲティー 果物，ソフトドリンク	七夕，横浜 マクドナルド
STATION	小田原，埼玉	区，不二家，アパート

の内容語抽出の問題点を示す．表 4 を見ると，FOOD クラスとして誤抽出された「焼き」や，STATION クラスとして抽出された「アパート」のように，内容語クラスの単語として不適当なものが抽出されていることもわかる．これら 2 単語がどのようなコンテキスト中で抽出されたかを確認した．抽出されたコンテキストを見ると，「焼き」の場合は，「明石 焼き に し て 食べよ」というコンテキストの「焼き」で抽出されていた．これは，「明石焼き」が FOOD クラスとして抽出されていれば問題ない．これは形態素解析によって，一単語となるべき単語が分割されてしまったために生じた問題と言える．また，学習データ中の STATION が出現するコンテキストを見ると，「名古屋 に 住ん で い る」のような「STATION に 住ん」というコンテキストが多く存在した．そのため，「アパート に 住ん で」というコンテキストで出現した「アパート」が STATION として抽出されたと考えられる．これは，今回 ME model での分類に用いた特徴量からでは判断できない．誤抽出単語を減らす特徴量の見直しが必要である．

5. 音声認識実験

5.1 実験条件

提案手法によって作成したクラス付き大規模コーパスから，クラス N-gram モデルを構築することで，音声認識率が向上することを確認する．今回は，7 種類のクラス N-gram モデルの音声認識率を比較した．この 7 種類のクラス N-gram モデルは，構築に使用したコーパスによって大きく次の 2 つに分けられる．

(1) **モデル** 1-X
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から構築したモデル．

(2) **モデル** 2-X
当該ドメインの記述文法と類似ドメインの大規模コーパスから構築したモデル．

さらに，クラス付与の方法によって，モデル 1-X，モデル 2-X は分けられる．

(1) **モデル** Y-1
類似ドメインの大規模コーパスには，クラスを付与していない．モデル 1-1, モデル 2-1 が該当する．モデル 1-1 の語彙サイズは 33178 単語，モデル 2-1 の語彙サイズは 33177 単語である．

(2) **モデル** Y-2
類似ドメインの大規模コーパスに対し，データベース中の内容語と完全に一致する単語へクラスを付与する．モデル 1-2, モデル 2-2 が該当する．語彙サイズは，モデル 1-2，モデル 2-2 ともに 33071 単語である．

(3) **モデル** Y-3
モデル Y-2 でのクラス付与に加え，表 3 の $n = 1$ のときに抽出した，未知の内容語 288 単語と完全に一致する単語へクラスを付与する．この 288 単語には，誤抽出単語も含まれる．つまり，このモデルは，現状の提案手法に人手を加えない場合のクラス N-gram モデルを示す．モデル 1-3 のみが該当する．モデル 1-3 の語彙サイズは 32806 単語である．

(4) **モデル** Y-4
モデル Y-2 でのクラス付与に加え，表 3 の $n = 1$ のときに抽出した未知の内容語のうち，人手で選別した正解単語 178 単語と完全に一致する単語へクラスを付与する．つまり，このモデルは，誤抽出を完全に除去した場合の提案手法の上限を示す．モデル 1-4, モデル 2-4 が該当する．モデル 1-4 の語彙サイズは 32902 単語，モデル 2-4 の語彙サイズは 32896 単語である．

なお，全てのモデルにおいて，当該ドメインの小規模コーパス・当該ドメインの記述文法には，それぞれ人手でクラスを付与している．

当該ドメインの小規模コーパスは，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．当該ドメインの記述文法は，上記と同様のドメインを想定し作成した記述文法からランダムで生成した 11,364 文を用いた．類似ドメイン大規模コーパスは，Yahoo!知恵袋の中カテゴリ「料理，グルメ，レシピ」の約 120 万文を用いた．当該ドメイン小規模コーパスまたは当該ドメインの記述文法と，類似ドメイン大規模コーパスをそのまま合わせると，当該ドメイン小規模な部分の影響がほとんど現れない．そこで，ここでは当該ドメイン小規模コーパス中の各文は 1 万回，当該ドメインの記述文法中の各文は 100 回複製し，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て，N-gram モデルを構築した．

上記で述べた比較する言語モデルをまとめたものを表 5 に示す．表 5 中の「小規模」は当該ドメインの小規模コーパスを，「記述文法」は当該ドメインの記述文法を，「大規

表 5 比較する言語モデル

該当モデル名	コーパスへのクラス付与		類似ドメインの大規模コーパスへのクラス付与に使用した単語		
	小規模 または 記述文法	大規模	検索対象データベースの単語	未知の内容語 (全て使用)	未知の内容語 (人手で除去)
モデル 1-1，モデル 2-1	○	–	–	–	–
モデル 1-2，モデル 2-2	○	○	○	–	–
モデル 1-3	○	○	○	○	–
モデル 1-4，モデル 2-4	○	○	○	–	○

模」は類似ドメインの大規模コーパスをそれぞれ表す．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から構築したクラス N-gram モデルが，該当モデル名のモデル 1-X で表されるモデルである．当該ドメインの記述文法と類似ドメインの大規模コーパスから構築したクラス N-gram モデルが，該当モデル名のモデル 2-X で表されるモデルである．○が付いている箇所が，言語モデル構築に使用したものを表す．

表 5 の見方を，モデル 1-3 を例として示す．モデル 1-3 は，「コーパスへのクラス付与」の「小規模」と「大規模」の部分に○が付いており，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にクラスが付与されていることを示す．また，「類似ドメインの大規模コーパスのクラス付与に使用した単語」の「検索対象データベースの単語」と「未知の内容語 (全て使用)」の部分に○が付いている．これは，類似ドメインの大規模コーパスに対して，検索対象データベースの単語と未知の内容語と完全に一致する単語へクラスを付与していることを示す．

クラスは，表 2 の内容語クラス 4 種類を設定した．各クラスの語彙サイズは，表 2 の登録単語数と一致する．各言語モデルのクラス内単語は，検索対象データベースの単語のみを使用した．また，クラス c_i に属する内容語 w_i のクラス内確率 $P(w_i|c_i)$ は各 c_i ごとに等確率とした．形態素解析器には 4 章の実験と同じく Mecab を使用した．

評価データには，言語モデルの構築とは別に，文献 [6] 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，未知の内容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 4480 発話 (14554 単語，うち内容語 1454) を使用した．評価指標には評価データ中の全単語と内容語それぞれに対する単語正解率，単語正解精度を用いた．音声認識には Julius[*3] を使用した．

5.2 言語モデルごとの音声認識率の比較

各言語モデルの音声認識率を比較し，提案手法によって音声認識率が向上することを示す．表 6 に，構築した言語モデルの音声認識結果を示す．ただし，表 6 中のモデル「closed」は，評価に用いた当該ドメインの 4480 発話の書き起こしから構築した単語 N-gram モデルを表す．closed は，この評価実験における音声認識率の上限の目安である．

表 6 構築した言語モデルの評価

言語モデル	全単語		内容語	
	Corr[%]	Acc[%]	Corr[%]	Acc[%]
closed	84.49	78.79	89.0	75.4
モデル 1-1	70.07	64.22	78.2	73.5
モデル 1-2	69.88	64.04	78.3	73.9
モデル 1-3	64.69	58.84	82.5	65.8
モデル 1-4	70.27	64.39	82.0	74.7
モデル 2-1	70.91	63.19	76.8	72.2
モデル 2-2	71.16	63.61	75.9	71.3
モデル 2-4	71.68	64.11	79.0	72.4

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から構築したモデル 1-X 間で比較する．表 6 において，内容語に対する評価を見ると，モデル 1-4 の Corr がモデル 1-1，モデル 1-2 に比べ，4%程度高いことがわかる．また，モデル 1-4 の Acc は，モデル 1-1，モデル 1-2 に比べ，1%程度高いことがわかる．この結果から，提案手法において，誤抽出単語を人手で除去した場合には，音声認識率は向上する．一方で，全単語に対する評価を見ると，モデル 1-1，モデル 1-2，モデル 1-4 には，ほとんど差異が見られないが，モデル 1-3 のみ，Corr，Acc ともに 5.5%程度低いことがわかる．この結果から，大規模コーパスへのクラス付与に誤抽出単語を用いると，音声認識率が下がることがわかる．今後は，誤抽出単語を減らすことが必要となる．

同様に，当該ドメインの記述文法と類似ドメインの大規模コーパスから構築したモデル 2-X 間で比較する．表 6 において，全単語・内容語それぞれに対する Corr，Acc を比較すると，モデル 2-1，モデル 2-2 に比べ，モデル 2-4 の結果が高いことがわかる．この結果から，記述文法を用いる場合でも，モデル 1-X 間での結果と同様，提案手法において，誤抽出単語を人手で除去した場合には，音声認識率は向上する．

以上の結果から，提案手法によって類似ドメインの大規模コーパスにクラスを付与したコーパスから言語モデルを構築することで，音声認識率の向上が期待できる．ただし，

[*3] http://julius.sourceforge.jp/

(1) 正解 :「うなぎ 消去」
(うなぎ : FOOD)
- モデル 1-1 :「何 消去」
- モデル 1-2 :「何 消去」
- モデル 1-4 :「うなぎ 消去」

(2) 正解 :「大須 の 居酒屋 を 教え て ください」
(大須 : LOCATION，居酒屋 : GENRE)
- モデル 1-1 :「大須 の 人 から を 教え て ください」
- モデル 1-2 :「大須 の 伸び た から を 教え て ください」
- モデル 1-4 :「大須 の 居酒屋 を 教え て ください」

(3) 正解 :「食べ物 ふぐ 削除」
(ふぐ : FOOD)
- モデル 1-1 :「食べ物 夫婦 削除」
- モデル 1-2 :「食べ物 夫婦 削除」
- モデル 1-4 :「食べ物 ふぐ 削除」

(4) 正解 :「ジャンル 和食 を 削除 し て ください」
(和食 : GENRE)
- モデル 1-1 :「ジャンル は チョコ を 削除 し て ください」
- モデル 1-2 :「ジャンル は チョコ を 削除 し て ください」
- モデル 1-4 :「ジャンル 和食 を 削除 し て ください」

図 2 認識結果の例

誤抽出単語にもクラスを付与した大規模コーパスを用いると，音声認識率が著しく下がるため，今後は，誤抽出単語の除去手法を検討する必要がある．

次に，提案手法において誤抽出単語を人手で除去した場合には，内容語が認識できるようになることを示す．音声認識結果の例を図 2 に示す．図 2 の (3) を見てみると，FOOD クラスに属する内容語「ふぐ」に対して，モデル 1-1，モデル 1-2 では「夫婦」と誤認識している．これに対して，モデル 1-4 では，正しく「ふぐ」を認識できている．また，(4) の文では，GENRE クラスに属する内容語「和食」に対して，モデル 1-1，モデル 1-2 では「は チョコ」と誤認識している．これに対して，モデル 1-4 では，正しく「和食」と認識できていることがわかる．これらの結果から，提案手法において誤抽出単語を人手で除いた場合に構築できるモデル 1-4 は，他の言語モデルに比べ内容語を正しく認識できていることがわかる．

6. おわりに

本稿では，ME model を用いた内容語抽出を行い，コーパスから未知の内容語が抽出できることを確認した．また，データベース中の内容語と，未知の内容語にクラスを付与したコーパスから，クラス N-gram モデルを構築し，音声認識率が向上することを確認した．ただし，未知の内容語は，人手で確認し，内容語として妥当であると判断された単語を用いた場合である．実験により，提案手法において誤抽出単語を人手で除去した場合には，提案手法により構築した言語モデルが，データベース中の内容語のみにクラスを付与したコーパスから作成したクラス N-gram モデルよりも，内容語が出現する発話パターンの音声認識率が向上することを確認した．

また，実験により明らかになった今後の課題を 4 つ挙げる．

(1) 誤抽出の抑制と除去．
(2) 形態素区切りの問題．
(3) 外部知識の利用．
(4) Bootstrap の実行．

誤抽出単語は，Bootstrap 式クラス付与・音声認識の両面で悪影響を及ぼすため，誤抽出単語を抑制・除去する必要がある．実験 1 の表 3 で示したように，現状の内容語抽出では，内容語として不適当な単語が数多く誤抽出される．誤抽出単語を正解クラスとみなして機械学習を行うと新たな誤抽出単語の増加を招く．また，実験 2 の表 6 で示したように，誤抽出単語にもクラス付与を行ったモデル 1-3 では，音声認識率が低下した．以上のような問題を引き起こすため，誤抽出を抑制する特徴量の再検討と，誤抽出単語の除去方法の検討が必要である．

内容語と形態素区切りが異なるという問題への対応が求められる．4.2 節で例に挙げた「明石 焼き」のように，内容語として抽出したい単語と形態素区切りが異なる場合がある．現在の手法では，内容語が複数の形態素に分割されてしまった場合には，正しく抽出できない．この問題は，内容語を抽出できないだけでなく，「焼き」を FOOD クラスとして抽出してしまうなど，誤抽出にもつながる．そのため，複数の形態素で一つの内容語になるような場合でも，正しく抽出することが求められる．

内容語辞書のエントリ数の増加や誤抽出の除去のために，外部知識の利用を検討する．現在は，大規模コーパスから未知の内容語を抽出し，内容語辞書のエントリを増加させている．これと並行して，他の百科事典的な知識を用いて辞書のエントリを増加させるという方法が考えられる．また，抽出した単語が誤抽出単語か否かの判断にも，外部知識の利用が考えられる．

Boostrap を複数サイクル回し，内容語抽出の様子を確認する必要がある．今回の実験では，Bootstrap を 1 サイクルしか回していない．そのため，複数回サイクルを回すことで，誤抽出による影響がどの程度波及するのか，何回程度まで回しても未知の内容語を抽出できるのかを確認する必要がある．

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．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] 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.

[3] 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.

[4] Adam L. Berger, Vincent J. Della Pietra, and Stephen A. Della Pietra. A maximum entropy approach to natural language processing. *Comput. Linguist.*, Vol. 22, No. 1, pp. 39–71, 1996.

[5] 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.

[6] 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-SLP-90, No. 20, pp. 1–7, 2012-01-27.