

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

駒谷 和範[†] 河原 達也[†] 清田 陽司^{††} 黒橋 禎夫^{††} Pascale Fung^{†††}

[†] 京都大学 情報学研究科

^{††} 東京大学 情報理工学系研究科

^{†††} Hong Kong University of Science and Technology, Weniwen Technologies Inc.

E-mail: [†]{komatani,kawahara}@kuis.kyoto-u.ac.jp, ^{††}{kiyota,kuro}@kc.t.u-tokyo.ac.jp,
^{†††}pascale@mail.weniwen.com

あらまし ドメインに適応した統計的言語モデルと構文情報を考慮したマッチングを用いて、音声入力により東京地区のレストランを検索するシステムを作成した。レストラン検索タスクにおける想定質問文から言語モデルを作成し、これを大規模コーパスから学習した統計的言語モデルと混合することにより、ドメインに適応した柔軟な言語モデルを作成する。自然言語処理部は想定質問文を構文解析したものを知識ベースとして保持し、音声認識結果も構文解析することで構文構造まで考慮したマッチングを行い、該当するレストランの Web サイトのインデックスをユーザに出力する。本システムはポータビリティを備えたシステム構成で柔軟な情報検索を実現している。

キーワード 音声認識, 検索システム, ドメイン適応, マッチング, 構文解析, 統計的言語モデル

Restaurant Search System with Speech Interface using Flexible Language Model and Matching

Kazunori KOMATANI[†], Tatsuya KAWAHARA[†], Yoji KIYOTA^{††}, Sadao KUROHASHI^{††},
and Pascale FUNG^{†††}

[†] Graduate School of Informatics, Kyoto University

^{††} Graduate School of Information Science and Technology, University of Tokyo

^{†††} Hong Kong University of Science and Technology, Weniwen Technologies Inc.

E-mail: [†]{komatani,kawahara}@kuis.kyoto-u.ac.jp, ^{††}{kiyota,kuro}@kc.t.u-tokyo.ac.jp,
^{†††}pascale@mail.weniwen.com

Abstract We present a flexible restaurant search system in Tokyo area with speech interface, which is characterized by domain-adapted statistical language model and matching method considering syntactic structures. We make an in-domain language model using sample sentences for restaurant queries, which is merged with the general language model trained on large corpora. The NLP module parses the speech recognition result and matches it with sample query sentences that are maintained as knowledge base. As the result of matching, indexes to corresponding Web restaurants' sites are listed on the browser. Our system realizes flexible responses for various utterances and portability to various tasks.

Key words speech recognition, search system, domain adaptation, matching method, syntactic structure, statistical language model

1. はじめに

音声認識技術の向上を受けて、個人向けの計算機の入力インタフェースとして音声が目ざされている。音声インタフェースを用いたアプリケーションの一つとして、音声による情報検索システムが挙げられる [1]。自然言語発話を入力として検索を行うためには、フロントエンドである音声認識とバックエンドである検索システムの双方が柔軟である必要がある。

音声対話システムでは、ドメイン固有の語彙や表現を認識できることがタスクの遂行に重要であるため、文法で言語制約や意味解釈規則が記述されることが多い。本研究室でも、データベース検索タスクにおいて、語彙や文法規則をデータベースから半自動的に抽出・生成するプラットフォームを作成した [6]。しかし文法規則に基づくシステムでは、ユーザの発話をシステムが認識可能な範囲にうまく誘導できない場合には、システムの想定外発話への対処が困難となるという問題があった [8]。

一方、近年大規模コーパスの整備を背景として、統計的言語モデルに基づく音声認識が成果を挙げている。統計的言語モデルでは多様な発話に対して柔軟に認識を行えるが、各ドメインごとに十分な学習データを確保することが困難であることが多い。また一般的な統計的言語モデルを用いる場合では、タスク遂行に重要なドメイン固有の表現や語彙（地名などの固有名詞）を正しく認識することは困難である。音声認識結果を Web の検索エンジンの入力として与えるシステム [1] においても、言語モデルをドメインに適応させない限り、十分なカバレッジ、認識精度は得られない。

音声認識結果を解釈する自然言語処理部も、音声認識部に統計的言語モデルを用いる場合には、あらゆる入力を処理できる必要がある。入力文から内容語を抽出し、タスクスロットを埋めるといった従来の音声対話システムでは、内容語を抽出し解釈するための意味解釈規則が記述される。VoiceXML による音声対話システム [4] もこの一つであり、タスクドメインに関する知識や解釈規則・対話管理規則を明示的に記述できる。しかしルールに基づいて解釈を行う場合には、文法による音声認識と同様に、想定外の入力に対して

適切な対処を行うことができない。

本稿では、ドメインに適した統計的言語モデルと柔軟なマッチング手法に基づくレストラン検索システムを提案する。大量の話し言葉のテキストから学習した一般的な統計的言語モデルと、少量のドメイン文から作成した統計的言語モデルを混合することにより、ドメインに適した柔軟な言語モデルを作成し、音声認識を行う。これにより、ドメイン特有の表現や語彙を認識可能としながら、多様な発話を柔軟に扱うことが可能である。自然言語処理部は、キーワードを意味スロットに埋めていくようなマッチングではなく、自然言語どうしのマッチングを行う [2]。構文情報まで考慮した柔軟なマッチングを行うため、あらゆる入力文に対してマッチングの尤度順に、結果を出力することができる。このように、本稿で提案するシステムは、様々な自然言語入力に対して柔軟に認識・解釈を行うことが可能である。さらにドメインの知識を自然言語で記述し、特殊な記述言語を用いる必要がないため、ポータビリティも良い。

2. システムの概要

本システムは、レストランに関するユーザの自由発話を入力とし、該当するレストランの Web サイトのインデックスを出力する。

自由発話の認識には、統計的言語モデルに基づく音声認識エンジン Julius [3] を用いる。言語モデルには、大量のテキストから学習した一般的な統計的言語モデルと少量のドメイン文から作成した統計的言語モデルを混合したものをを用いる (3. 章)。ドメイン固有のレストラン名や地名は、ドメイン文から作成した言語モデルを混合することにより認識可能となる。多様な発話に対応するために必要な、話し言葉一般の言語制約は、大量のテキストデータから学習したモデルから得られる。

音声認識結果はそのまま自然言語処理部に渡される (4. 章)。自然言語処理部では、保持している知識ベースに基づき、入力されたユーザの要求に近いものを計算し、出力する。知識ベースとして、レストランの各データに対応する想定質問文を構文解析したものが蓄えられており、入力された音声認識結果を構文解析したものとマッチングをとることにより、類似度が計算され

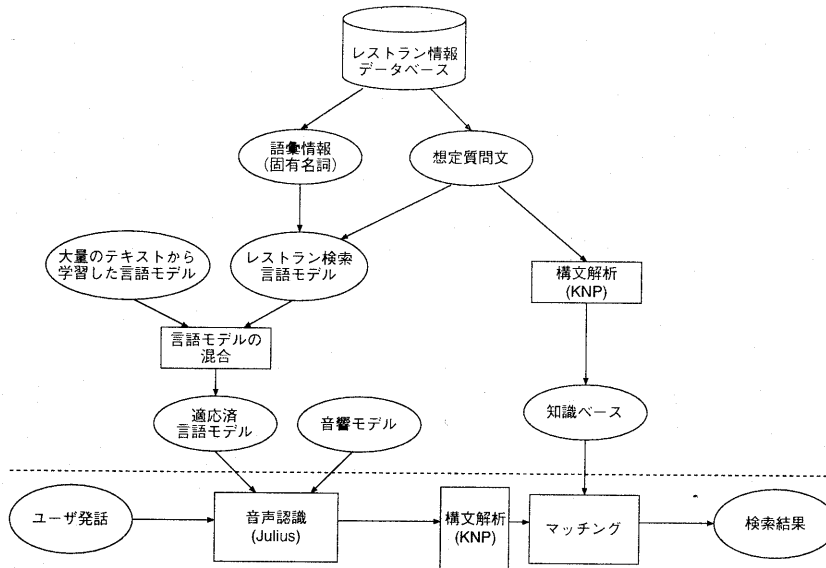


図1 システムの処理の概要

る．得られた各部の類似度に基づいてマッチングの確信度を計算し，確信度の大きいものを検索結果として順に出力する．出力されたレストラン名は，そのレストランのWebサイトへのポイントとなっているため，ブラウザ上でレストラン名をクリックすることにより，ユーザはより詳しい情報を得ることができる．

3. ドメインに適応した柔軟な言語モデル

該当ドメインにおける小規模のテキストから作成した言語モデルと，大量のテキストコーパスより学習した一般的な言語モデルを混合することにより，ドメインに適応した柔軟な言語モデルを作成する．

該当ドメインのテキストとして，東京地区のレストラン検索に用いる想定質問文を用いる．これらは人手により記述され，図2に挙げるようなクラスを用いた想定質問文335文と，各レストランに対応する「インターネットで予約できるレストランは？」のような想定質問文131文からなる．これらのテキストから，クラスN-gramモデルを作成した．クラスは図3に挙げる7種類である．

得られたクラスN-gramモデルの単語辞書に，各クラスに対応する固有名詞を登録することにより，レストラン検索タスクの言語モデルとし

<RESTAURANT>はどこですか
 <RESTAURANT>の電話番号を教えてください
 <RESTAURANT>の営業時間を教えてください
 <district>の<RESTAURANT>はどこですか
 <RESTAURANT>に<TRANSPORT>で行きたいのですが
 <PLACE>で一番有名な<FOOD_TYPE>の店はありますか
 おいしい<SNACK>はどこで食べられるんですか
 <PLACE>の近くのおいしい<REST.TYPE>を教えてください

図2 クラスを用いた想定質問文の例

た．この言語モデルの語彙サイズは925である．固有名詞を単語辞書に登録する際には，形態素解析プログラムによって得られた読みを人手により修正した．

ここで作成した言語モデルは疑似的なクラスN-gramモデルである．すなわちクラス内のすべての単語の生起確率が1であるものに相当する．

このようにして作成したレストラン検索言語モデルと，グルメレシビドメインの話し言葉言語モデル（語彙サイズ19447）[9]を，N-gramモデル融合ツール[10]を用いて混合する．混合比は1:1とした．混合の結果得られた言語モデルの語彙数は20370である．グルメレシビドメインの言語モデルは大量の話し言葉テキストから学習さ

<RESTAURANT>: レストラン名 (245)
 <FOOD.TYPE>: 料理のタイプ (68)
 - 中国料理, 日本料理, イタリア料理, ……
 <PLACE>: 所在地 (125)
 - 恵比寿, 新宿, 日比谷, 有楽町, ……
 <REST.TYPE>: レストランのタイプ (7)
 - 居酒屋, 喫茶店, ラーメン屋, ……
 <SNACK>: 料理名 (129)
 - お好み焼き, パエリア, 回転寿司, ……
 <TRANSPORT>: 交通手段 (5)
 - タクシー, 電車, バス, JR, 地下鉄
 <district>: 区 (4)
 - 中央区, 港区, 渋谷区, 新宿区

図3 固有名詞のクラスとその内容

れたもので、多様な入力に対する柔軟な対処が可能であるが、レストラン検索タスクに特有の表現や固有名詞が含まれていない。少量のドメイン文から作成した言語モデルと混合することにより、対象ドメインに特有な表現や固有名詞を認識可能で、かつ柔軟な言語モデルとなる。

4. 音声認識結果と知識ベースの柔軟なマッチング

4.1 レストランサイトの収集と想定質問文の作成

東京都心の4区(中央区, 港区, 渋谷区, 新宿区)に所在する245店のレストランのWebサイトを人手で収集した。さらに、各店のWebサイトに対して、それぞれ複数の想定質問文を作成した。想定質問文の例を以下に示す。

- 八重洲大飯店への行き方を教えてください。
- 道玄坂でフランス料理はどこで食べられますか?

- 宴会に利用できる居酒屋は?
- 芸能人がよく来るラーメン屋は?

4.2 ユーザ質問文と想定質問文のマッチングシステム

すべての想定質問文を、あらかじめKNP[7]により構文解析しておく。音声認識されたユーザの質問も、同様にKNPにより構文解析する。マッチングシステムは、想定質問文の構文木とユーザ質問文の構文木とのマッチングを、[11]のアルゴリズムにしたがって行う(図4)。

まず、探索範囲の絞り込みのために、想定質問

文に対して、ユーザ質問文中の自立語が何語含まれるかを評価値として予備選択を行う。

次に、予備選択で取り出された想定質問文の構文木とユーザ質問文の構文木との類似度計算を行う。まず、ユーザ質問文の構文木中の各文節に対して、想定質問文の構文木の最も類似する文節を探す。文節間の類似度は以下の基準により計算する。

- 自立語の一致(文節内で1語目) 4点
- 自立語の一致(文節内で2語目以降) 1点
- 係りタイプ(表層格など)の一致 1点

さらに、ユーザ質問文の構文木中の各文節の類似点を合計し、それを以下のように一致点の満点で正規化し、想定質問文の確信度とする(一致点の満点は同一構文木との間の文節類似点の総和)。確信度は次式により定義する。

$$\frac{\text{文節類似点の総和} \times \sqrt{\text{文節類似点の総和}}}{\text{ユーザ質問文の満点} \times \sqrt{\text{想定質問文の満点}}} \quad (1)$$

最後に、それぞれのレストランについて、対応する想定質問文の確信度の最大値をそのレストランの確信度とする。このマッチングの結果、確信度が最大のレストランと、確信度が最大値の90%以上のレストランを、確信度の大きい順に並べて出力する。

5. 実装と動作例

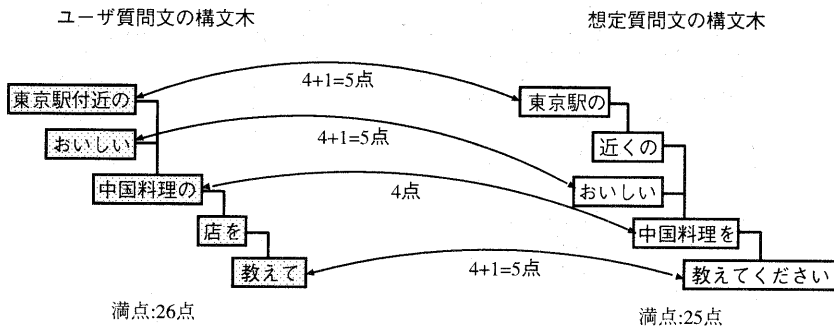
5.1 システムのポータビリティ

本システムを他のドメインに移植する際には、言語モデルと知識ベースの再構築が必要となる。この作業の大部分を既存のツール(KNP[7]・言語モデル作成ツール[9])を用いて行うことができる。本システムを構築する際に、システム開発者が行った手作業は以下の2点である。

(1) Webページを収集し、それに対応する想定質問文を作成する

(2) 固有名詞に対する読みを訂正する

各Webサイトに対応する想定質問文の大部分は、図2で示したテンプレートに固有名詞を代入することにより自動的に生成することができる。これらの想定質問文の作成は人手で行う必要があるが、特殊な記述言語を用いる必要はなく、そのまま自然言語で記述することができる。この自然言語で記述された想定質問文は、言語モデルの作成と知識ベースの作成の両方で用いられ、両方



$$\text{確信度} = \frac{(5+5+4+5) \times \sqrt{5+5+4+5}}{26 \times \sqrt{25}} \times 100 = 64 (\%)$$

図4 ユーザ質問文と想定質問文の構文木マッチング

のモジュール（音声認識部・自然言語処理部）のドメイン適応に役立つ（図1）。

従来の多くの音声対話情報検索システムでは、入力文を解釈するにはアプリケーション特有の意味スロットを用意していた[8]。これに対して本システムでは、各Webサイトに対応する想定質問文を自然言語で記述すればよいため、システム開発者の負担は少なく、新たなWebサイトをシステムに加えるのも容易である。

5.2 動作例

前章までで述べたシステムを、ノートPC（Linux）とワークステーション（Solaris）の両方に実装し、動作を確認した。

出力にはWebブラウザを用い、入力した発話の音声認識結果とマッチングの結果得られたレストラン名が表示される（図5）。出力されたレストラン名には対応するレストランのWebサイトへのリンクが張っており、マウスでクリックすることにより、Webサイトから詳細な情報を得ることができる。ルールに基づくシステムとは異なり、記述した想定質問文に存在しない以下のようなユーザの質問に対しても、柔軟に認識・解釈を行い、妥当な検索結果を出力することができる。

- 新宿でいちばんの焼肉屋を教えてください
- 港区にあるお洒落なバーを教えてください
- すき焼きのおいしい店はありますか
- おいしいビールが飲みたい

問題点として、誤認識により誤った単語が音声認識部から出力された場合に、誤った部分に対応する候補も同時に出力されてしまうことが挙げ

られる。現在は音声認識部と自然言語処理部が全く独立に動作し、検索結果を出力しているが、音声認識における信頼度[5]やマッチングの尤度を利用することにより、確からしくない候補をユーザに知らせたりすることにより、対話的に誤りを解消する方法を検討する必要がある。

6. まとめ

本稿では、ドメインに適応した統計的言語モデルを作成し、構文情報を用いた自然言語どうしのマッチングを行うことにより、任意の表現の音声認識・解釈可能なレストラン検索システムを作成した。音声認識部及び自然言語処理部は、単純なルールに基づくものではないため、様々な自然言語音声入力に対する柔軟な対処が可能であり、想定質問文として記述されていないユーザの質問に対しても、妥当な検索結果を出力できることを確認した。

本システムのドメイン適応のための手作業は、想定質問文の記述と固有名詞に対する読みの訂正の2種類であり、残りの作業は既存のツールを利用して行える。想定質問文の記述に特殊な記述言語は必要なく、自然言語を用いるためシステムの移植性・拡張性も高い。このため種々のドメインの情報検索へ適用可能である。

謝辞

音声認識用言語モデルの作成には、連続音声認識コンソーシアム^(注1)により配布されているモ

(注1) : <http://www.lang.astem.or.jp/CSRC/>

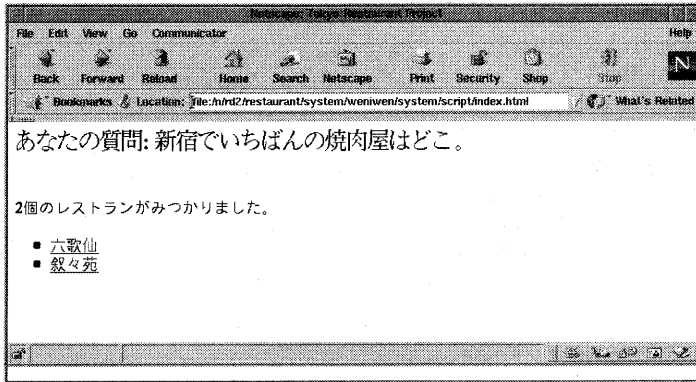


図5 動作画面

デルとツールを使用した。

文 献

- [1] <http://www.sharp.co.jp/liquiy/liquiy02.html>.
- [2] Kurohashi, S. and Higasa, W.: Dialogue Helpsystem based on Flexible Matching of User Query with Natural Language Knowledge Base., *Proc. of 1st ACL SIGdial Workshop on Discourse and Dialogue*, pp. 141-149 (2000).
- [3] 李晃伸, 河原達也, 堂下修司: 単語トレリスインデックスを用いた段階的探索による大語彙連続音声認識, 電子情報通信学会論文誌, Vol. J82-DII, No. 1, pp. 1-9 (1999).
- [4] 植田喜代志, 秋田祥史, 荒木雅弘, 西本卓也, 新美康永: VoiceXMLのマルチモーダル化の検討, 情報学研報, 2001-SLP-38-7 (2001).
- [5] 駒谷和範, 河原達也: 音声認識結果の信頼度を用いた頑健な混合主導対話の実現法, 情報処理学会研究報告, SLP-30-9 (2000).
- [6] 田中克明, 河原達也, 堂下修司: 汎用的な情報検索音声対話プラットフォーム, 情報処理学会研究報告, 98-SLP-24-14 (1998).
- [7] 黒橋禎夫, 長尾真: 並列構造の検出に基づく長い日本語文の構文解析, 自然言語処理, Vol. 1, No. 1, pp. 35-57 (1994).
- [8] 安達史博, 駒谷和範, 河原達也: 音声対話情報検索システムにおける想定外の発話の分析とその対処, 人工知能学会研究会資料, SIG-SLUD-A001-2 (2000).
- [9] 河原達也, 住吉貴志, 李晃伸, 武田一哉, 三村正人, 伊藤彰則, 伊藤克巨, 鹿野清宏: 連続音声認識コンソーシアム2000年度版ソフトウェアの概要と評価, 情報学研報, 2001-SLP-38-6 (2001).
- [10] 長友健太郎, 西村竜一, 小松久美子, 黒田由香, 李晃伸, 猿渡洋, 鹿野清宏: 相補的バックオフを用いた言語モデル融合ツールの構築, 情報処理学会研究報告, 2001-SLP-35-9 (2001).
- [11] 日笠亘, 藤井綱貴, 黒橋禎夫: 入力質問と知識表現の柔軟なマッチングによる対話的ヘルプシステムの構築, 情報処理学会研究報告, 99-NLP-134-14 (1999).