

音声対話システムにおけるユーザ適応技術の統合手法の提案

守時 理裕

荒木 雅弘

京都工芸繊維大学情報工学専攻

概要

近年では、音声認識や自然言語処理の発展に伴い、音声対話システムにおける機械学習を用いた様々なユーザ適応手法が研究されてきた。しかし、従来の音声対話システムにおける学習では、音響モデルや言語モデルといった個別要素に関しては各々に適した機械学習手法が提案されてきたが、これら個別要素における適応技術を統合したユーザ適応のためのシステムは研究されてこなかった。本研究では、これらのユーザ適応技術を統合することにより、音声対話システム全体におけるユーザ適応を行い、さらに、統合により生じる、あるレベルの学習が他のレベルの学習に与えるような平行学習問題を考慮した、様々なレベルにおける適応機能を実現する手法を提案する。

1 はじめに

近年の音声認識・自然言語処理における統計的手法の発展を受け、音声対話システムにおいて機械学習を用いた様々なユーザ適応手法が研究・開発されてきた。例えば、音響モデルの適応では、MLLR法が話者適応の基本的な手法として利用されている [1]。言語モデルにおいては、大規模コーパスで学習したモデルを併用したタスクに依存した少量のコーパスからの言語モデルの学習 [2] や、被験者のモチベーションを上げるため active learning を利用した効率的な学習用データの選別 [3] などがある。また、対話管理に関しては、マルコフ決定過程を用いた強化学習による対話戦略の獲得 [4] などがある。

このように、音声対話システムにおける学習の問題は、個別要素に関してはそれぞれ適した機械学習手法などの話者適応技術が進み、ある程度の学習が可能である。しかし、これらの技術を統合し、様々なレベルで適切に振る舞うような音声対話システムはこれまで研究されてこなかった。本研究では、これらのユーザ適応技術の統合に加え、その統合の際に生じる、あるレベルの学習が他のレベルの学習に与える並行学習問題を解決するような、様々なレベルにおける適応機能を実現する手法を提案する。

2 音声対話システムにおけるユーザ適応技術

本研究で扱う話者適応は、音声対話システムにおける以下に示す各モジュールで行う。

- 言語モデル
対話ログから失敗事例を抽出、ディクテーションを行い、確信度計算結果の高いものをタスク依存言語モデルに反映させる。
- 音響モデル
MLLRを用いて、対話の進行に伴って収集される音声ログを適応用データとして動的に音響モデルに反映させる。
- 概念モデル
音声対話システムにおける知識を体系化させたオントロジーを用いて、認識結果への概念タグ付けおよび概念タグ系列から意味表現への変換規則を対話コーパスから学習する [5]。
- 対話管理
マルコフ決定過程 (MDP) を用いて、対話コーパスから対話戦略の強化学習、状態の分割・集約を動的に行う。

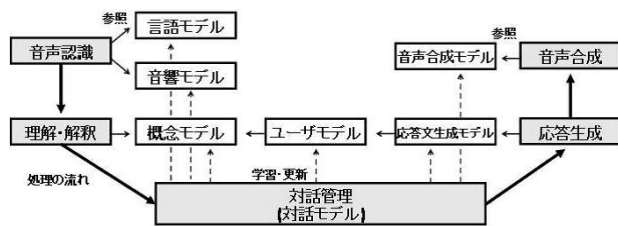


図 1: 音声対話システム構成図

- 応答生成

コーパスを利用した統計的な応答生成を行う。

- 音声合成

ユーザに応じた話速やポーズ長のコントロールを行う。

音声対話システム全体の構成を図 1 に示す。本稿では、この 6 つのモジュールのうちの言語モデルの学習と対話戦略の学習の間で生じる並行学習問題について検討する。

3 適応技術の統合により生じる影響

本研究は、これまで行われてきた複数の適応技術を単に結合するだけのものではない。家電操作を行うような家庭用ロボットなどにおける音声インタフェースを想定し、少数特定話者への適応を、音響から対話までの様々なレベルで行う。その際、多レベルでの学習を平行して行えば、互いのレベル間において様々な影響を及ぼし合うことが考えられる。多レベルでの学習を平行して行うことにより、得られる長所と短所が考えられる。

3.1 統合による知識の共有

多レベルでの平行学習の利点としては、少量のデータでは適応が難しいとされてきた言語モデルや対話戦略、応答生成の適応に対して、多レベルでの知識の共有が可能になると考えられる。以下にその 3 点を挙げる。

- 意味解析規則の知識を言語モデルの学習に利用
- 対話戦略の学習を意味解析規則の学習に利用

- 言語モデルの知識を応答生成の学習に利用

まず、概念モデルにおける意味解析規則の学習が言語モデルに与える影響として、言語モデルの学習において音声対話システムにおける知識である意味レベルのオントロジーを用いれば、学習コーパスに出現しないような単語においても、ユーザのドメインレベルを獲得することでオントロジーから部分集合を計算し、そのユーザが使う可能性の高い単語に対する言語モデル重みを調整、修正することが可能であると考えられる。

また、対話戦略の学習が概念モデルの意味解析規則の学習に与える影響も考えられる。対話管理部で得られるユーザとシステムとの間の対話の主導権情報をもとに、システム主導であれば意味解析において単純な文に、ユーザ主導であれば複雑な文に重みを加えることも可能であると考えられる。

また、言語モデルの学習が応答生成の学習に与える影響も考えられる。対話においてユーザが頻繁に使用する単語を応答に用いることで、ユーザの知識レベルに応じた応答が可能になると考えられる。

3.2 統合による平行学習問題

しかし上記で挙げた利点とは対照的に、多レベルでの学習を平行して行うことの欠点も考えられる。その一つとして、対話戦略の学習が言語モデルの学習に与える影響が考えられる。言語モデルの学習は対話の各状態において行われるが、対話戦略の学習において状態が新しく生成・併合されれば、その時点までの言語モデルの学習が無駄になる可能性がある。このような問題を、本稿では「動的に評価基準が変わる環境下での平行学習問題」と呼ぶ。

本稿では、そのデメリットの一つとして考えられる、対話戦略の学習が言語モデルの学習に与える問題に注目する。

4 動的な環境下での平行学習問題

4.1 対話戦略の学習に伴う対話状態の遷移・分割

音声対話システムにおける話者適応では、言語モデルの学習は対話の各状態において行われるが、対話戦略の学習において状態が新しく生成されたり併合したりすると、それまでの言語モデルの学習が無駄になってしまう。つまり、対話状態が遷移の際に併合する場合は、その遷移確率で各言語モデルに重み付けをして混合させればよいが、対話戦略が学習され状態が分割された場合、それまで学習してきた言語モデルを、分割された各々の対話状態でどのようにして持たせるのかという問題が生じる。その様子を図2に示す。ここで、言語モデル LM1 を持つ対話状態が、対話戦略の学習により分割された各対話状態に持たせる言語モデルを、LM1' 及び LM1'' とする。

一般的な解決法としては、状態が遷移する毎にその対話状態までのコーパスを用いて逐次的に言語モデルを再学習する手法が考えられるが、対話状態が分割を繰り返す毎に、LM1 用のデータを LM1', LM1'' 用にラベル付けして分割する必要がある。

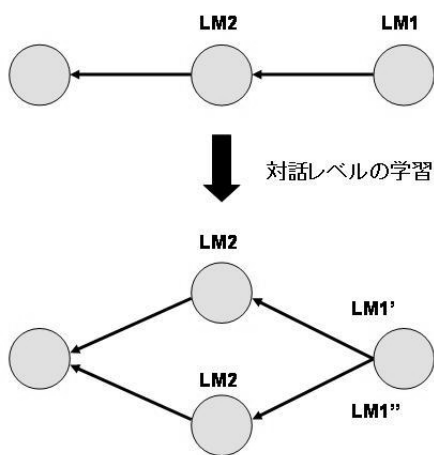


図 2: 対話戦略の学習に伴う対話状態の分割

4.2 EM アルゴリズムを用いた言語モデルの適応

そこで本研究では、LM1 用コーパスに対するラベリング、及び分割された対話状態に対して持たせる

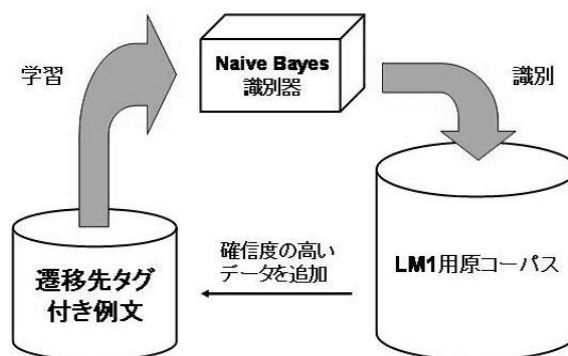


図 3: EM アルゴリズムを用いたコーパスの分類

言語モデルの作成を試みる。その様子を図3に示す。

まず、それぞれの遷移先に対応する例文を幾つか用意し、そこから初期識別器を得る。識別器には Naive-Bayes アルゴリズムを用いる。次にその識別器を用いて LM1 用原コーパスの各データにラベル付けを行い、確信度の高いものを例文集合に加え、それを用いて識別器を学習することを繰り返す。データ数が増加しなくなった時点でタグに基づいてコーパスを LM1', LM1'' 用に分類し、各データを用いて各々の対話状態に言語モデル LM1 を適応させる。このことにより、対話の各遷移・分割ごとの全対話履歴を用いた言語モデルの再学習にかかる計算量は大幅に削減できると考えられる。

5 結論

音声対話システムにおける話者適応技術の統合により生じる影響について述べてきた。本研究では、その中でも特に、技術の統合の際に学習に関して悪影響を及ぼす可能性のある対話戦略の学習が、言語モデルの学習に与える問題についてその解決手法を提案した。今後は、この手法をもとに音声対話システムの実装、学習を行い、実験及びその解決手法の評価を行いたい。

参考文献

- [1] C. Leggetter and P. Woodland(1995): Flexible Speaker Adaptation using Maximum Likelihood Linear Regression; In: Proceedings of

the ARPA Spoken Language Technology Workshop, S. 104-109

- [2] He Wei, Li Honglian and Yuan Baozong(2003): Domain Adaptation Augmented By State-Dependence In Spoken Dialogue Systems; Eurospeech 2003 Vol.2, pp/883-886
- [3] G. Tur, R. E. Schapire and D. Hakkani-Tur(2003): Active Learning for Spoken Language Understanding; In Proc ICASSP, Vol.1, PP.274-279
- [4] S. Singh, M. Kearns, D.Litman and M. Walker(1999): Reinforcement Learning for Spoken Dialogue Systems; In Advances in Neural Processing Systems, Vol.12
- [5] M. Araki, Y. Kimura, T. Nishimoto, and Y. Niimi(2001): Development of a machine learnable discourse tagging tool; In Proc. of 2nd SIG-DIAL Workshop on Discourse and Dialogue, pp.20-25,2001