

強化学習に基づく逐次環境クラスタリングによる音声認識

Speech Recognition Using Incremental Environment Clustering based on Reinforcement Learning

西田 昌史 堀内 靖雄 市川 熹
Masafumi Nishida Yasuo Horiuchi Akira Ichikawa

千葉大学大学院自然科学研究科
Graduate School of Science and Technology, Chiba University

This paper describes a novel approach based on online unsupervised adaptation and clustering using reinforcement learning. Reinforcement learning technique is a computational approach to learning whereby an agent tries to maximize the total amount of reward it receives when interacting with a complex, uncertain environment. The adaptation progresses based on rewards that represents correctness of outputs. The adapted models gradually accumulate and cluster with the environmental conditions and can immediately adapt by selecting the optimal model from the clusters. We conducted speech recognition experiments by a connected digit recognition in noisy environments including the variation of speakers and noises. The results verify that the proposed method has a higher recognition performance than the conventional adaptation method.

1. はじめに

音声対話など実環境において頑健な音声認識を実現するためには、背景雑音や話者の変化に対応する必要がある。例えば、カーナビではエアコンや音楽を使用しているときや、天候や交通量などで背景雑音に変化し、使用する話者も変化する。このような環境では、雑音や話者の変化を逐次的に適応し環境モデルのクラスタリング処理が有効であると考えられる。

これまで我々は、雑音や話者の変動を逐次的に適応する手法として、強化学習を用いた教師なし環境適応法を提案した [1]。本手法は、事前に環境データやモデルなどの知識を必要とせず、オンラインで環境の変化に適応することが可能であった。しかし、従来の強化学習による手法では逐次的に環境への適応を行うため、過去に適応したモデルを利用できなかった。また、直前の適応モデルに基づいて学習を行うため、即座に環境に適応することが困難であった。そこで、本研究では環境の変化に応じて適応モデルを蓄積し、オンラインでクラスタリングを行う手法を提案する。本手法により、クラスタから最適なモデルを選択して強化学習を行うことで、即座に環境に適応することができると考えられる。また、自動的に適応モデルを随時蓄積することが可能となる。

2. 強化学習に基づく適応とクラスタリング

2.1 Actor-Critic 法

本研究では、強化学習の一手法である actor-critic 法 [2] を用いる。本手法の概念図を図 1 に示す。actor は、状態から行動への確率的政策に従って行動を実行する。critic は、actor の政策下で実行された行動に対して、新しい状態の価値を評価し、その結果が期待されたものより良かったかどうかを判断する。その評価結果は、式 (1) に示す TD(Temporal Difference) 誤差 δ_t の形式で actor に与えられる。

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \quad (1)$$

ここで、 r_{t+1} は行動 a_t により状態 s_t から s_{t+1} に移ったときに得られた報酬、 $V(s_t)$ は状態 s_t の価値関数を示している。

連絡先: 西田 昌史, 千葉大学大学院自然科学研究科, 〒263-8522 千葉市稲毛区弥生町 1-33, nishida@faculty.chiba-u.jp

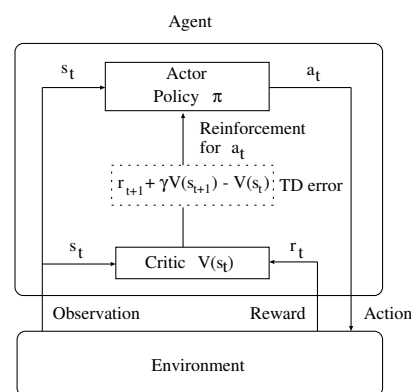


図 1: actor-critic 法の概念図

actor は TD 誤差が正の場合、実行した行動の選択確率を高め、負の場合実行した行動の選択確率を下げることで政策を改善する。また、critic はそれに伴い状態の価値を更新する。

2.2 音響モデルの適応

音響モデルの適応は、MAP 推定に基づいて HMM の各状態に属する全混合分布の平均ベクトルに対して行う [3]。事前分布が正規分布 $N(\mu_0, \sigma_0)$ に従うと仮定すると、混合分布の平均ベクトル $\hat{\mu}$ は次式のように推定される。

$$\hat{\mu} = \frac{\lambda\mu_0 + T\bar{x}}{\lambda + T} \quad (2)$$

ここで、 \bar{x} は適応データ、 T は適応データのフレーム数、 λ は更新係数を表している。

本研究では、MAP 推定における更新係数 λ の制御を政策とみなして学習することにした。また、連続数字列の認識を行うため、入力と出力は数字列となっている。強化学習における出力の正しさとしては、フレーム毎に最尤の HMM の状態をもつ数字列と発話全体での最尤な数字列を比較し、これらの数字列の一致度を基準とした。行動 a_t を実行して状態 s_{t+1} に遷移した際の状態価値 $V(s_{t+1})$ としては、数字列の一致度の割合とし、報酬 r_{t+1} は数字列の一致度が前発話に比べて高ければ 1.0、低ければ 0.5 の値を与えた。この状態価値と報酬によ

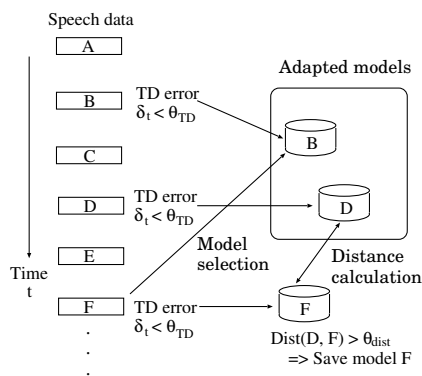


図 2: actor-critic 法に基づくモデル選択とクラスタリング

り TD 誤差 δ_t を求めて、式 (3) により政策の改善を行う。比較した数字列が一致した区間は信頼度が高いとして更新係数を小さくして適応を行う。また、数字列が一致しなかった区間は信頼度が低いとして、更新係数を大きくして適応を行う。

$$\lambda \leftarrow \lambda \pm \beta \delta_t \quad (3)$$

2.3 適応モデルの選択とクラスタリング

actor-critic 法に基づくモデル選択とクラスタリング手法の概念図を図 2 に示す。本手法は、TD 誤差が閾値より小さくなれば、環境が変化したとみなして適応モデルの選択を行う。モデルの選択は、入力データと現在のモデルならびにクラスタのモデルとそれぞれ距離を計算し、距離が最小となる適応モデルを選択し強化学習を行う。入力データとモデルとの距離は、HMM の各状態 s の全混合分布の平均ベクトルから重心ベクトル $\bar{\mu}_s = 1/M \sum_{i=1}^M \mu_{si}$ を求め、入力データと HMM の各状態の重心ベクトルとのユークリッド距離を用いた。

その際、現在の適応モデルとクラスタの適応モデルとの距離を計算し、最小距離が閾値より大きければ現在のモデルはクラスタに蓄積される。モデル間の距離は、HMM の各状態の重心ベクトルを束ねたベクトル $\bar{\mu}_h = (\bar{\mu}_{h1}, \bar{\mu}_{h2}, \dots, \bar{\mu}_{hS})$ 間のユークリッド距離に基づき、式 (4) により計算した。

$$Dist(A, B) = \sum_{h=1}^H \|\bar{\mu}_h^A - \bar{\mu}_h^B\|^2 \quad (4)$$

ここで、 H は HMM の総数である。

3. 評価実験

3.1 実験条件

本研究では、雑音環境下連続数字認識コーパスである AURORA-2J を用いて、適応を行わず初期モデルでの認識、従来手法である MAP 適応、提案手法による音声認識実験を行った。また、MAP 推定や actor-critic 法におけるパラメータは、それぞれ予備実験により設定した。

音声データの特徴量には、12 次 MFCC、パワーのそれぞれの 1 次、2 次回帰の計 39 次元のパラメータを用いた。音響モデルは、16 状態、20 混合分布の HMM により各数字をモデル化した。初期モデルは、110 名の話者による 8440 発話のクリーンな音声により学習を行った。

評価には、8 種類の雑音 (Subway, Babble, Car, Exhibition, Restaurant, Street, Airport, Station) を 3 種類の SN 比 (10,

15, 20 dB) で重畳させた音声データを用いた。評価データは、話者 45 名による約 430 発話で、発話を 3 つに分割しそれぞれ異なる雑音を重畳させたもの 16 パターンを用いた。なお、3 つに分割した最初と最後の部分は同じ雑音を用いた。

3.2 実験結果と考察

各 SN 比に対する認識結果を表 1 に示す。actor-critic 法に

表 1: 各 SN 比に対する音声認識結果 (%)

| | ベースライン | MAP | 強化学習 | 強化学習 + クラスタリング |
|-------|--------|------|------|-------------------|
| 20 dB | 90.5 | 97.1 | 97.2 | 97.6 |
| 15 dB | 66.0 | 91.3 | 90.9 | 90.7 |
| 10 dB | 42.6 | 68.0 | 69.4 | 70.1 |

よるモデル適応は、ベースラインに比べて 20dB において 7%、15dB において 27%、10dB において 39% の認識精度の改善が得られた。また、従来の MAP 適応に比べて 10dB において認識精度が 1.4% 向上した。したがって、従来手法では、認識結果や環境の変動に関係なく適応の重みが一定であるのに対して、強化学習により環境の変動に応じて適応の重みを制御することが有効であることがわかった。

actor-critic 法によるモデル選択とクラスタリング手法は、10dB において従来の MAP 適応に比べて 2.1%、actor-critic 法によるモデル適応と比べて 0.7% 認識精度が向上した。従来の MAP 適応や強化学習による適応手法では、直前の環境で適応したモデルを逐次的に学習していくため、環境が大きく変化した場合現在の環境に即座に適応することができない。それに対して、提案手法では環境の変化に応じて過去の適応モデルから最適なものを選択して強化学習を行うことで、話者や雑音の変動をより頑健に適応することができたと考えられる。

4. おわりに

本研究では、強化学習の一手法である actor-critic 法による環境適応とクラスタリングに基づく音声認識手法を提案した。本手法は、環境の変化に応じて最適な適応モデルを選択し強化学習を行うことで、即座に環境へ適応することができる。また、オンラインで自動的に環境の変化に応じた適応モデルを蓄積することができる。AURORA-2J を用いた音声認識による評価実験を行った結果、提案手法は話者や雑音の変化に応じて動的にモデルを選択し適応することで、従来の適応手法に比べて高い認識精度を得ることができた。

謝辞

本研究では、IPSJ SIG-SLP 雑音下音声認識評価ワーキンググループの AURORA-2J データベースを用いた。

参考文献

- [1] 西田 昌史, 堀内 靖雄, 市川 薫 “強化学習を用いたオンライン話者・環境適応に基づく音声認識,” 日本音響学会 2004 年秋季研究発表会, 2-4-10, pp.105-106, 2004.
- [2] R. S. Sutton and A. G. Barto, “Reinforcement Learning: An Introduction,” MIT Press, 1998.
- [3] J. L. Gauvain and C. H. Lee, “Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains,” IEEE Trans. Speech and Audio Processing, vol. 2, no. 2, pp. 291-298, 1994.