

遺伝的アルゴリズムを用いた 構音障害者の音声特徴量抽出に最適なランダム行列の生成*

☆片岡悠一郎, 中鹿亘, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

構音障害とは言語障害の一部をなすもので, 内的言語能力自体は正常で声も出るが, 正しく聴き取れる発音ができない症状をいう. 発話が不安定で手足も不自由な場合など, 音声に頼るしかない状況が考えられ, 言語障害者を対象とした音声認識の実現が必要である. 構音障害の原因は種々あるが, その一つに脳性麻痺が挙げられる. 彼らは意図的な動作時や緊張状態にある場合に筋肉の制御が難しくなり, アテトーゼと呼ばれる不随意運動を伴う. アテトーゼ型の構音障害者の発話スタイルは健常者と大きく異なり, 認識精度が著しく低下する. しかし, 構音障害者などの言語障害を患う方を対象とした音声認識の研究は非常に少ないのが現状である. そこで, 本研究では, アテトーゼ型の脳性麻痺による構音障害者を対象とした音声認識の実現を目指している.

著者らは, これまでに構音障害者の音声特徴量抽出にランダムプロジェクションを適用し, 有効性を確認してきた. ランダムプロジェクションとは, 元の特徴量空間における任意の2点間のユークリッド距離が, 写像先の特徴量空間においても高い確率で保存される, という性質を持つ空間写像の一手法で, その変換写像行列の各要素がある確率分布に従うランダムな値として定義される点に特徴を持つ. ランダムプロジェクションでは, 様々な(無限の)ランダム写像行列を生成することが出来るため, 文献 [1] では, ROVER 法による結果統合を行っていた.

本稿では, 遺伝的アルゴリズムによる最適なランダム行列の選択方法を検討する. 遺伝的アルゴリズムとは, 生命の進化の過程をもとに作られたアルゴリズムであり, 交叉, 突然変異, 選択の操作により, 次世代へ優秀な個体を生成し, これを繰り返すことにより, 最適解を探索するアルゴリズムである.

2 ランダムプロジェクション

ランダムプロジェクションは, n 次元ユークリッド空間から k 次元ユークリッド空間へランダムに写像する空間写像の手法である [2]~[7]. ある n 次元の元特徴量ベクトル x が与えられたとき, k 次元 ($k \leq n$) の変換後の特徴量ベクトル y は次のように表される.

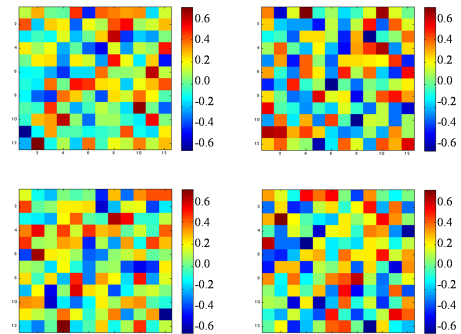


Fig. 1 Examples of Random Matrices

$$y = Rx \quad (1)$$

ここで R は $n \times k$ の写像行列である. 写像行列 R は確率的にある値をとる行列として定義されるが, R の各要素が $N(0,1)$ に従うランダムな値からなるとき, 任意の2点間の距離が高い確率で $(1 \pm \epsilon)$ に収まることが証明されている ($0 \leq \epsilon \leq 1$) [2, 3]. 本稿では次のようなルールでランダム写像行列 R を作成する.

- 標準正規分布 $N(0,1)$ に従う要素を持つ $n \times k$ の行列 R を作成
- Gram-Schmidt の直交化手法を用いて R を直交化
- 行ベクトルを大きさ1で正規化

上記の生成方法により, ランダム写像行列 R は標準正規分布 $N(0,1)$ から無限に生成することが出来る. Fig. 1 に本稿で用いたランダム写像行列 R の例を示す.

3 モデル適応

ランダムプロジェクションを行う度に音響モデルを学習して認識率を算出すると, 計算コストが高くなるという問題がある. そこで, 本章では, ランダムプロジェクションによって得られた特徴量ごとに音響モデルを学習するのではなく, あらかじめランダムプロ

* Generation of random matrices using a genetic algorithm for dysarthria speech feature extraction, by Yuichiro Kataoka, Toru Nakashika, Tetsuya Takiguchi, Yasuo Ariki (Graduate School of System Informatics, Kobe University)

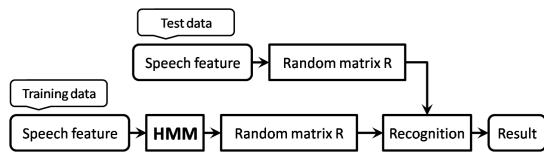


Fig. 2 Overview of Random Transformation on features and HMM

ジェクションを行う前の音声特徴量で音響モデルを学習しておき、その音響モデルに対してランダム写像行列を用いた線形変換を行う。あらかじめ学習された音響モデルの平均値ベクトルと分散共分散行列に対して、ランダム写像行列を用いた線形変換を行うことで、複数回音響モデルを学習することなく、ランダムプロジェクション特徴量での認識を可能にする。

Fig. 2 に音声特徴量と音響モデルの変換の流れ図を示す。あらかじめ変換前の学習データを用いて音響モデルを学習する。そして、認識を行うデータに対してランダムプロジェクションを行い、ランダムプロジェクション特徴量を得ると同時に、学習された音響モデルに対しても同様のランダム写像行列で線形変換を行い、得られた音響モデルで認識を行う。

以下に、音響モデルに HMM を用いた際のモデル適応の手法について述べる。HMM の出力確率は、多次元混合正規分布で表現されているため、各々の正規分布の平均値ベクトルと分散共分散行列に対してランダムプロジェクションを行う。多次元正規分布の平均値ベクトル μ_x と分散共分散行列 Σ_x は特徴量ベクトル x_i を用いて式 (2)、式 (3) のように表現される。

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

$$\Sigma_x = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(x_i - \mu_x)^T \quad (3)$$

N は、特徴量ベクトル x_i のデータ数である。ここで、特徴量ベクトル x_i が式 (1) を用いて $y_i = Rx_i$ と変換されたとき、平均値ベクトルと分散共分散行列は次のように変換できる。

$$\begin{aligned} \mu_y &= \frac{1}{N} \sum_{i=1}^N Rx_i \\ &= R\mu_x \end{aligned} \quad (4)$$

$$\begin{aligned} \Sigma_y &= \frac{1}{N} \sum_{i=1}^N (Rx_i - \mu_y)(Rx_i - \mu_y)^T \\ &= R\Sigma_x R^T \end{aligned} \quad (5)$$

このように音響モデルに対してランダムプロジェクションを行うことで、特徴量ごとに音響モデルを学習するコストを省き、効率良く有効な音声認識を行うことが可能となる。

4 遺伝的アルゴリズムを用いたランダム行列の生成

遺伝的アルゴリズムは、自然界における生物・生体の集団遺伝に基づく進化の過程を模倣したアルゴリズムであり、交叉、突然変異、選択の操作により、次世代へ優秀な個体を生成し、これを繰り返すことにより、最適解を探索するアルゴリズムである [8]。本稿では、以下のアルゴリズムを用いた。

1. ランダム行列を n 個生成し、第 1 世代の個体とする。
2. 前章で述べた、モデル適応の手法を用いて音声認識実験を行い、それぞれのランダム行列に対して認識率を求め、その数値を各個体の評価値とする。
3. 現世代の中で最も認識率の高い個体を次世代の 1 番目の個体とする。
4. 交叉、突然変異、選択のいずれかを、予め定めておいた確率で行い、次世代の 2～ n 番目の個体とする。
5. 次世代の個体を現世代の個体とし、2～4. を第 k 世代まで繰り返す。
6. 第 k 世代の中で最も優秀な個体が、求めるランダム行列である。

Fig. 3 に、以上のアルゴリズムの流れを示す。次に、交叉・突然変異・選択の説明をする。

交叉 (crossing-over)

現世代の中から、予め定めておいた確率で 2 つの個体を選択する。選択された 2 つの個体の遺伝子のいずれかを、予め定めておいた確率で選択し、次世代の個体の遺伝子 (gene) とする。ここで、遺伝子とは、ランダム行列の各行ベクトルのことである。この操作を遺伝子の数だけ繰り返し、それらを集めて生成された行列に Gram-Schmidt の直交化と正規化を行って得られた行列を、次世代の個体とする。

突然変異 (mutation)

新しいランダム行列を生成し、次世代の個体とする。

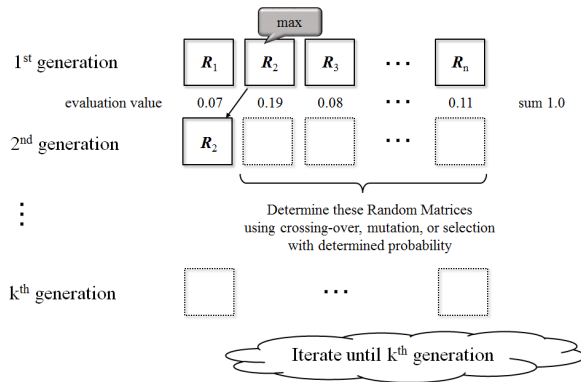


Fig. 3 Overview of Genetic Algorithm

選択 (selection)

予め定めておいた確率で現世代の個体を選択し、次世代の個体とする。

Fig. 4 に、交叉・突然変異・選択の流れを示す。

本稿では、上記の遺伝的アルゴリズムにより、構音障害者の音声認識に最適なランダム行列を生成することにより、音声認識率を向上させる手法を提案する。

5 評価実験

提案手法の有効性を示すため、評価実験を行う。

構音障害者 2 名を対象とした孤立単語認識実験を行い、実験データセットとして、構音障害者 2 名による ATR 音素バランス 216 単語 × 4 回発話を用いる。ただし、これらのデータには欠落データが存在する。音声の標準化周波数は 16kHz であり、音響モデルは monophone-HMM で、各 HMM の状態数は 5、状態あたりの混合分布数は 8 である。ここでは、第 1 発話を評価データ、第 2~4 発話を学習データとして用いる。

今回の実験における、遺伝的アルゴリズムの各パラメータは以下の通りである。

- 個体数 $n = 100$
- 世代数 $k = 200$
- 「交叉」と「選択」で選ばれる個体の確率は、個体の評価値の合計を 1 になるように標準化したもの
- 交叉における遺伝子の選択確率は、選択した 2 つの個体のいずれも 0.5
- MFCC[12] to RP[12] + Δ MFCC[12] : 12 次元 MFCC 特徴量を 12×12 のランダム写像行列で変換し、 Δ 特徴量を組み合わせた 24 次元特徴量

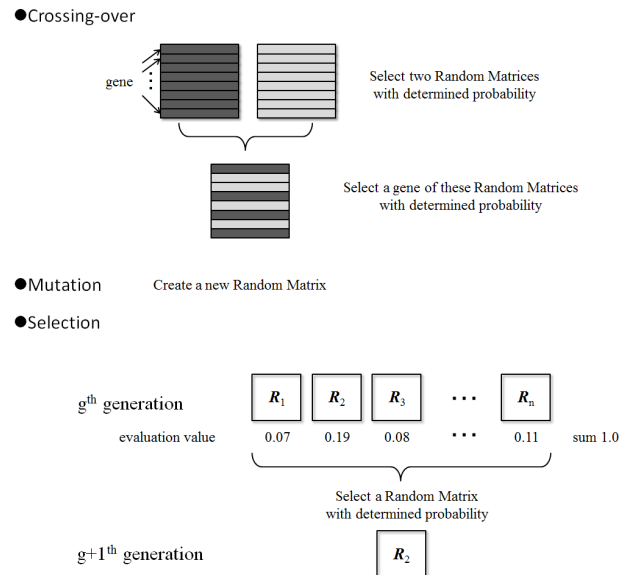


Fig. 4 Overview of crossing-over, mutation, and selection

Table 1 Recognition rates [%] using a genetic algorithm for the begining generation and the last generation

	begining generation	last generation
speakerA	63.81	67.14
speakerB	93.98	94.91

(交叉、突然変異、選択) のいずれかを選択する確率を調整し、最も音声認識率が向上したものを実験結果とする。

6 実験結果

Table 1 に構音障害者 2 名について、遺伝的アルゴリズムの最初の世代の中で最も高い認識率と、最終世代の中で最も高い認識率の比較を示し、Fig. 5 に世代毎の認識率の推移のグラフを示す。

以上の結果より、話者 A については 3.33%、話者 B については 0.93% の認識率の向上が見られた。2 話者共に認識率の向上が見られ、提案手法の有効性が示された。話者 A が話者 B と比較して認識率向上率が大きかったのは、話者 A の方が、ベースラインの認識率が低かったためであると考えられる。

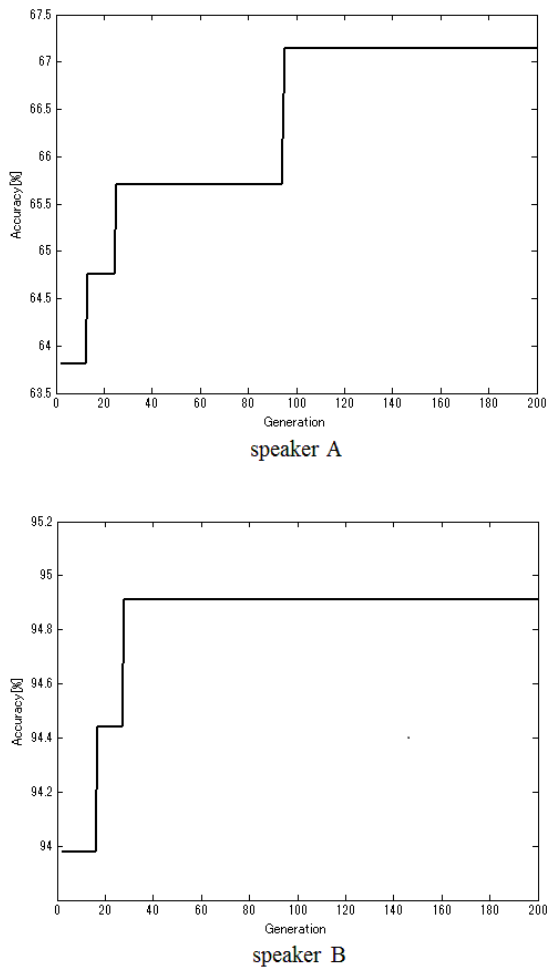


Fig. 5 Recognition results

7 おわりに

本稿では、遺伝的アルゴリズムを用いた構音障害者の音声認識精度を向上させる手法を提案した。遺伝的アルゴリズムにより、認識に最適なランダム行列を生成することにより、音声認識精度が向上した。実験結果より、遺伝的アルゴリズムを用いて音声認識精度を向上させる手法は有効であることが示された。今後は、遺伝的アルゴリズムの様々なパラメータを調整して認識精度向上について検討する。

参考文献

- [1] T. Takiguchi, M. Yoshii, and Y. Ariki, “ACOUSTIC MODEL TRANSFORMATIONS BASED ON RANDOM PROJECTIONS”, in ICASSP, pp. 1933-1936, 2012.
- [2] W. B. Johnson and J. Lindenstrauss, “Extensions of Lipshitz Mapping into Hilbert Space,” in Conference modern analysis and probability, volumn 26 of Contemporary Mathematics, pp. 189-206, 1984.
- [3] D. Achlioptas, “Database-friendly random projections,” In Proc. ACM Symp. on the Principles of Database Systems, pp. 274-281, 2001.
- [4] S. Dasgupta and A. Gupta, “An elementary proof of the Johnson-Lindenstrauss lemma,” Technical Report TR-99-006, International Computer Science Institute, Berkeley, California, USA, 1999.
- [5] S. Dasgupta, “Experiments with random projection,” in Uncertainty in Artificial Intelligence, pp. 143-151, 2000.
- [6] R. I. Arriaga and S. Vempala, “An algorithmic theory of learning: robust concepts and random projection,” Proc. IEEE Symposium on Foundations of Computer Science, pp. 616-623, 1999.
- [7] E. Bingham and H. Mannila, “Random projection in dimensionality reduction: Applications to image and text data,” Proc. 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2001), pp. 245-250, 2001.
- [8] 貝原俊也編著, “オペレーションズリサーチ”, オーム社, pp. 93-94, 2004-12.