

潜在特徴の自己組織化による 時系列を考慮した状態空間の可視化

吉田 洋生[†] 長尾 智晴^{††}

[†] 横浜国立大学 理工学部

^{††} 横浜国立大学 大学院環境情報研究院

1. はじめに

近年, 知能情報学の分野ではニューラルネットワーク (Neural Network; NN) を利用した手法の発展が目覚ましい。しかし, NN の推論根拠は一般に不透明である。そのため, 運用段階や意思決定に責任を伴う現場での用途は限定的であり, 解釈性を有するモデルが求められている。そこで本稿では, NN で学習した潜在特徴量を低次元に埋め込むことで, データがもつ状態空間を可視化する手法を提案する。本手法によって, データに潜在する状態の時系列遷移が視覚的に表現でき, 医療現場における病態の予見・診断などに応用できると考えられる。

2. 提案手法

本稿では, 変分オートエンコーダー (Variational Auto-encoder; VAE)[1] を用いてデータの潜在特徴量を学習し, その潜在特徴量に自己組織化を適用することで離散的な状態空間を獲得する手法を提案する。

本手法は SOM-VAE[2] と呼ばれる従来手法を参考にし, 入力画像を再構築する過程で, 潜在特徴量を離散的空間に埋め込む中間処理を施す。このとき, 離散的空間が自己組織化マップ (Self-Organizing Maps; SOM)[3] の特性を持ち, かつ空間上の距離が状態のステップに比例するように, 埋め込みを学習する。これにより獲得した空間は, 状態がクラスタとして, かつその遷移が時系列に基づくように表現された状態空間となる。これを可視化することで, 時系列的な予見の獲得を図る。

3. 実験設定

本実験では, 手書き数字の画像データセット MNIST を用いて, その状態遷移の可視化を行った。0 から 9 までの数字を状態とみなし, さらに隣り合う 2 つの数字を 5 等分に線形補間して中間状態を生成することでデータセットを拡張し, 入力に用いた。

また, 可視化した状態空間のクラスタリング性能を, 正規化相互情報量 (Normalized Mutual Information; NMI) と純度 (Purity) を用いて定量的に評価した。

表 1. 提案手法と従来手法[2]の NMI と Purity

	NMI	Purity
従来手法	0.594	0.731
提案手法	0.418	0.473

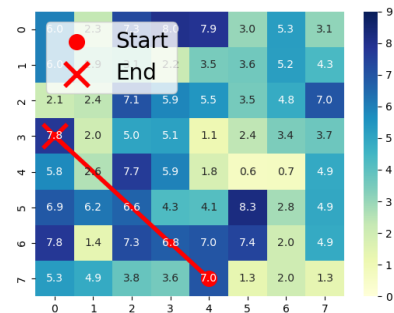


図 1. 獲得した状態空間と数字 7→8 における遷移

4. 実験結果

提案手法と従来手法[2]の MNIST データセットに対する NMI と Purity を表 1 に示す。どちらも従来手法に及ばず, クラスタリング性能は低下する結果となった。これは, 提案手法で導入した空間上の距離が状態のステップに比例するような学習が, 状態の分離とトレードオフの関係にあるためだと考えられる。

また図 1 に提案手法により可視化した状態空間を示す。状態ラベルの平均を表す数字とその色付けから隣接するセルの状態が近いことが見てとれ, 直感的な解釈性を持つ空間を学習できている。さらに, 数字 7 から 7.25, 7.5, 7.75 と遷移したとき, その延長線上の 8 に最も近いセルに終点がプロットされており, 次の状態が予測できることを示唆している。一方で, 遷移の途中に乖離した状態をまたいでいる点, 1 ステップの遷移幅が一定でない点に関しては改善の余地がある。

5. まとめ

本稿では, VAE で学習した潜在特徴量を自己組織化し, 状態空間として可視化する手法を提案した。また, 線形補間で拡張した MNIST データセットを用いて, 実際に状態遷移を獲得できることを確認した。

今後は, データセットの対象範囲を広げ, 実際の医用データセットを用いて手法の有効性を検証するとともに, クラスタリング精度の向上を図り, 状態の時系列的な予見などの医療診断への応用を検討する。

参考文献

- [1] Kingma, Diederik P et al. "Auto-Encoding Variational Bayes.", *ICLR*, 2014.
- [2] Fortuin, Vincent et al. "Deep self-organization: Interpretable discrete representation learning on time series.", *ICLR*, 2019.
- [3] Kohonen Teuvo. "Self-organized formation of topologically correct feature maps", *Biological Cybernetics*, 1982.