

量子機械学習に基づく画像解析

理学専攻・情報科学コース 1940638 朝岡 日向子

1 はじめに

画像解析を行うための機械学習の過程において、次世代計算機の一つであるイジングマシンを導入する。イジングマシンは、量子力学の理論に着想を得て開発された組合せ最適化問題を解くことに特化した計算機だが、機械学習問題に応用することも可能である [1, 2]。本研究では、富士通デジタルアニーラというイジングマシンを利用する [3]。行列分解のアルゴリズムで辞書学習を行い、その過程の分解の計算を最適化問題として扱うことによって、マシンを利用する計算を可能とする。辞書学習で解析する対象は人間の顔画像である。行列分解で顔の特徴を抽出し、それらを用いて分解前の画像の再現や、人物ごとの分類を行った結果を示す。今回は2種類の辞書学習のアルゴリズムを導入し、それぞれの結果を比較する。

2 モデル

2.1 Nonnegative/Binary Matrix Factorization

二値制約非負値行列分解 (Nonnegative/Binary Matrix Factorization; NBMF)[4] は、行列を非負値行列とバイナリ行列に分解することによって、低次元化して表現する手法である。画像の画素値から作成した行列にこの手法を適用すると、解となる2つの行列は画像の特徴とその選び方の組合せを表すことになる。それらを利用して、元の画像よりも次元やデータ量が削減された解析が可能となる。

はじめに、画像 m 枚の画素値から、行列 V を作成する。 V の列ベクトルがそれぞれの画像にあたり、画像1枚の画素数が n であるとすると、 $n \times m$ の行列になる。この行列に NBMF を施すと、 $n \times k$ の非負値行列 W と、 $k \times m$ のバイナリ行列 H に分解される。

$$V \approx WH \quad (1)$$

このとき、 $W_{ij} \geq 0$, $H_{ij} \in \{0, 1\}$ が常に成り立つ。式 (1) を満たすように、 W と H に関して最小二乗法を交互に利用して逐次更新する。

$$W := \arg \min_{X \in \mathbb{R}^{n \times k}} \|V - XH\|_F + \alpha \|X\|_F \quad (2)$$

$$H := \arg \min_{X \in \{0, 1\}^{k \times m}} \|V - WX\|_F \quad (3)$$

ここで、 $\|\cdot\|_F$ はフロベニウス距離を表す。行列 W は実数行列であるため、式 (2) は射影勾配法 [5] を用いて更新する。一方、行列 H は二値行列であるため、アニーリングマシンで更新式を解くことができる。 H の更新は列ごとに行う。

$$H_l = \arg \min_{q \in \{0, 1\}^k} \|V_l - Wq\|_2 \quad (4)$$

ここで、 H_l , V_l はそれぞれ行列 H , V の第 l 列である。本研究で利用するデジタルアニーラは、Quadratic

unconstrained binary optimization (QUBO) 模型を入力として受け付けるため、式 (4) を式 (5) の QUBO 表現に変換する。

$$f(\mathbf{q}) = \sum_i a_i q_i + \sum_{i < j} b_{ij} q_i q_j \quad (5)$$

$$a_i = \sum_r W_{ri}(W_{ri} - 2V_{ri}), \quad b_{ij} = 2 \sum_r W_{ri} W_{rj}. \quad (6)$$

すると、評価関数 (5) を最小にする $\mathbf{q} \in \{0, 1\}^k$ が、解として返ってくる。これが行列 H の各列にあたるため、式 (4) の計算を m 回行うことによって、 H の全体を更新する。1回の更新後と更新前の W 同士のフロベニウス距離が 10^{-4} 以下になったら収束したと見なし、逐次更新を終了する。

また、NBMF の類似手法として、非負値行列分解 (Nonnegative Matrix Factorization; NMF)[6] が知られている。同じく行列分解の手法であるが、行列 H が非負値の実数行列となり、 $W_{ij} \geq 0, H_{ij} \geq 0$ が常に成り立つ。

2.2 Method of Optimal Directions

Method of Optimal Directions (MOD)[7] は、画像から作成した行列を、特徴のパッチから成る辞書行列と、どのパッチを用いて画像を再現するかをスパースに表す行列に分解する。はじめに、画像を複数枚のパッチに切り分け、それぞれのパッチを各列ベクトルとして行列 Y を構成する。この Y を、非負値の辞書行列 D とバイナリ行列 X の2つに分解する。

$$Y \approx DX \quad (7)$$

式 (7) を満たすために、 D, X を逐次更新することによって以下の式 (8) の最小化を行う。

$$\min_{D, X} \{\|Y - DX\|_2^2\} \quad (8)$$

はじめに初期化した D を用いて、以下の更新式で X を列ごとに更新する。

$$\mathbf{x}_l = \arg \min_{\mathbf{x}_l} \{\|\mathbf{y}_l - D\mathbf{x}_l\|_2^2\} \quad (9)$$

式 (9) は、NBMF においてアニーリングマシンを用いて解いていた更新式とほぼ同じ形である。従って、式 (5) の QUBO に変換し、アニーリングマシンで解く。 X の各列について解いた後、更新した行列全体を用いて D を求める。

$$D = \arg \min_D \{\|Y - DX\|_2^2\} = YX^T (XX^T)^{-1} \quad (10)$$

$\|Y - DX\|_2^2$ が十分に小さな値になるまで、 D, X の逐次更新を繰り返す。

3 結果

3.1 使用する画像データについて

The Olivetti faces dataset (AT & T Laboratories Cambridge) から、訓練画像 120 枚、テスト画像 20 枚の人間の顔画像のデータセットを取得する。訓練画像を I_{train} 、テスト画像を I_{test} とする。今回の実験では 20 人分のデータを利用する。それぞれの人物は番号付けされており、その番号を「ラベル」と呼ぶ。

3.2 画像再現

I_{train} の訓練画像を画素値から成る行列 V に変換し、NBMF のアルゴリズムを用いて行列分解を行うことによって顔画像の特徴を抽出する。 V を構成する元画像と、それを再現した画像にあたる WH の構成ベクトルとの間で、二乗平均平方根誤差 (Root Mean Squared Error; RMSE) を計算する。この誤差が小さければ、正確な行列分解ができたと考えられる。全ての画像で誤差の算出を行い、全体での平均を行列分解の結果とする。今回は、 I_{train} を同一条件で 5 回分解した。NMF のアルゴリズムでも同様の分解を行い、その結果を比較する。

NBMF における最終的な RMSE が最も低くなった結果をグラフに出力したものが、図 1 である。

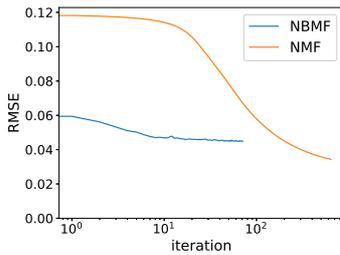


図 1: 各アルゴリズムが収束するまでの RMSE の推移

収束までの更新の平均回数は、NBMF の方が圧倒的に少ない。最終的な RMSE は NMF の方が小さいが、NBMF は 1 回目の更新から比較的小さい誤差になっており、即ち少ない更新回数で元画像に近いものを再現できる。これは、特徴を選択する際に最適化を行った利点であると考えられる。

3.3 画像分類

3.3.1 NBMF を用いる画像分類

NBMF のアルゴリズムを用いて画像进行分类する際に、2 種類の手法を用いる。

1 つ目は k 近傍法で、画像ベクトル間の距離を利用する手法である。この手法で、 I_{test} を先ほど学習した 5 回の訓練データを用いて判定したところ、平均正解率は 93% となった。また、NMF による行列分解で得られた特徴行列でも k 近傍法で分類を行ったところ、平均正解率は 89% となった。

2 つ目の手法では、Neural Network の構造を導入する。行列 V の下に、出力層としてラベルを表す行列をつけて分解すると、結果として得られる行列 W のうち、画素値である n 行目までは画像の特徴を持ち、それ以降はラベルの特徴を表すと考えられる。これらの特徴行列を用いて判定を行ったところ、平均正解率は 97% となった。

3.3.2 MOD を用いる画像分類

MOD では、 I_{test} を構成する画像を 1 枚ずつ行列 Y_{test} に変換し、 $Y_{\text{test}} \approx DX_{\text{test}}$ の近似を行う。 X_{test} は、辞書行列 D の中からどの特徴パッチを選んで Y_{test} の再現を試みたかをバイナリの値で表す行列になるため、 X の中に構成が似ている画像がある場合は、同じ様な特徴を選んでいて類似すると考えられる。この考え方に基づき k 近傍法で判定を行ったところ、平均正解率は 85% となった。

3.3.3 分類結果の比較

各手法で行った分類の結果を、表 1 で比較する。

表 1: 分類結果の比較

	NBMF		NMF	MOD
	k 近傍法	NN		
正解率	93%	97%	89%	85%
マシンでの計算回数	6480 回			320 回
次元削減率	$\frac{60}{1024}$ 倍			$\frac{256}{256}$ 倍
スパース度	41.7%	63.6%	39.9%	11.7%

NBMF の方が高精度な分類ができているが、MOD の方も十分に良い結果が出せている。計算回数やスパース度の面で考えると、MOD のアルゴリズムを選択するメリットもあると言える。ただし、次元を削減した行列表現を得たい場合は、NBMF を利用する必要がある。

4 まとめ

画像行列を分解し、解として得られる再現画像との誤差を検証することによって、行列分解を最適化問題として解くことの有用性を確認できた。また、分解によって抽出した画像の特徴を用いて、数種類の手法で分類を行った。どの手法でも高い正解率で分解できたが、それぞれ利点異なるため、目的によって使い分けることが可能である。

参考文献

- [1] T. Kadowaki and H. Nishimori, Phys. Rev. E **58**, 5355 (1998).
- [2] M. Benedetti, J. Realpe-Gómez, R. Biswas, and A. Perdomo-Ortiz, Phys. Rev. X **7**, 041052 (2017).
- [3] S. Matsubara, and et al., in *25th Asia and South Pacific Design Automation Conference*, 2020.
- [4] D. O'Malley, V. V. Vesselinov, B. S. Alexandrov, and L. B. Alexandrov, PLOS ONE **13**, e0206653 (2018).
- [5] C. J. Lin, Neural Computation **19**, 2756 (2007).
- [6] D. D. Lee and H. S. Seung, Nature **401**, 788 (1999).
- [7] K. Engan, S. O. Aase, and J. H. Husøy, Proc. ICASSP **5**, 2443 (1999).