

二値制約非負値行列分解を用いる画像分類

朝岡 日向子^{1,a)} 工藤 和恵^{1,2} 白井 達彦³ 田中 宗^{4,5} 戸川 望³

概要: 画像解析は、幅広い分野で利用されている技術であり、我々の生活の身近なところでも様々な場面で使われているが、多くのデータ量を要する。本研究では、画像を行列に変換し分解することによって、通常よりもメモリを削減して解析を行う技術について検討する。また、行列計算に次世代型の計算機を用いることによって、速度面での発展も期待する。行列分解のアルゴリズムとして、二値制約非負値行列分解を利用する。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。画像を行列とみなしてこの手法を適用すると、画像群が持つ特徴を得ることができる。また、計算過程のバイナリ行列を算出する箇所では、富士通デジタルアニーラを利用する。本論文では、二値制約非負値行列分解を用いて顔画像を解析する。実験として、画像群を分解して得られる特徴を用いて画像の分類を行い、その精度を検証する。

キーワード: 画像分類, 行列分解, アニーリング, 組合せ最適化

1. はじめに

本論文では、画像解析を行うための機械学習の過程において、イジングマシンという量子力学の理論に基づく、あるいはそれに着想を得た次世代計算機を導入する。イジングマシンの中でも量子アニーリング方式のものは、重ね合わせの理論に基づく構造を取り入れることによって、高速な計算を可能としている。すでに実用化されているものもあり、D-Wave マシンなどが知られている [1]。量子アニーリングは、組合せ最適化問題を解くためのアルゴリズムである [2–4]。組合せ最適化問題は、スケジューリング問題など様々な社会問題解決のために利用されているが、その計算は古典計算機だと膨大な時間を要する。計算にイジングマシンを用いることによって、問題解決の高速化を図る研究がなされている [5–10]。また、イジングマシンは典型的な最適化問題以外にも用いられ、機械学習問題に利用した結果も報告されている [11–16]。本研究では、画像解析に富士通デジタルアニーラ [10] というイジングマシンを利用する。デジタルアニーラは、シミュレーテッドアニーリング [17] 及びレプリカ交換法 [18] を用いて、イジング問

題を高速に解く。本研究では行列分解のアルゴリズムで機械学習を行い、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、デジタルアニーラを利用する計算を可能とする。

今回は、二値制約非負値行列分解 (Nonnegative/Binary Matrix Factorization; NBMF) [16] のアルゴリズムを利用する。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。計算過程のバイナリ行列を算出する箇所は、特徴選択の最適化問題となる。今回の実験では、最適化の計算においてデジタルアニーラを利用する。

本研究では、解析する対象として人間の顔画像を用いる。行列分解で顔の特徴を抽出し、それらを用いて人物ごとの分類を行った結果を示す。過去にも一部同じ内容を含む結果を発表したが、本論文では利用するデータ数を増やして実験を行った [19]。また、2種類の異なる分類の手法を提案する。K近傍法を用いる手法では、ベクトル間距離を利用して判定を行う。もう一方では、NBMFの行列分解のアルゴリズムに、Neural Networkの概念を導入する。Neural Networkとは、脳の神経回路を模した数理モデルを利用する機械学習手法の一種である。今回は、行列分解のアルゴ

¹ お茶の水女子大学

² 東北大学大学院

³ 早稲田大学 基幹理工学部

⁴ 慶應義塾大学 理工学部

⁵ 早稲田大学 グリーン・コンピューティング・システム研究機構

a) asaoka.hinako@is.ocha.ac.jp

リズムでこの手法の構造を再現し、画像分類に利用する。それぞれの結果で、結果にどのような違いが表れるか検証する。

本論文の構成は次のようになっている。第2章では、NBMFによる行列分解のアルゴリズムについて解説する。第3章では、実験で利用した画像データについて述べる。第4章では、第2章のアルゴリズムを画像解析に利用する際の手法とその結果を示す。第5章では、NBMFにNeural Networkの構造を導入し、画像分類を行う手法とその結果を示す。第6章では、アルゴリズムで設定したいいくつかのパラメータと、その値に設定した根拠について述べる。最後に、結果から得られる考察を第7章にて述べる。

2. Nonnegative/Binary Matrix Factorization

Nonnegative/Binary Matrix Factorization (NBMF) の行列分解アルゴリズムを用いて、画像が持つ特徴を抽出する手法について解説する。

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、 $n \times m$ の行列になる。この行列にNBMFを施すと、画像の特徴を表す $n \times k$ の非負値行列 W と、それぞれの画像がどの特徴を持っているのかを表す $k \times m$ 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。 W の各列は、画像が持つ特徴で、即ち基底となる画像にあたる。 W の列数 k は基底画像の数となり、これは任意の値に設定する。 H は V と同じく、列ベクトルが画像を表すが、 k 次元二値ベクトルになっているため、元の V よりもデータ量が削減されている。

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。フロベニウス距離の算出は、 A を実行列、その要素を a_{ij} とすると、式(4)のように定義される。

$$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2} \quad (4)$$

式(2)の α は正の値をとるハイパーパラメータで、行列 W をマシンに学習させるにあたり、挙動を制御し、過学習を防ぐためあらかじめ設定しておく必要のある値である。今

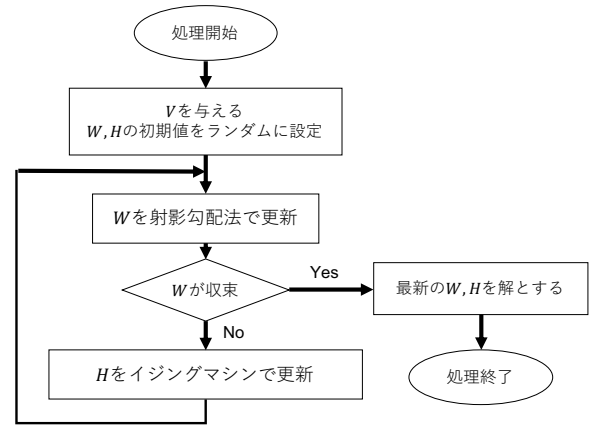


図1 NBMFのフローチャート

Fig. 1 A flowchart of NBMF algorithm.

回は、 $\alpha = 1.0 \times 10^{-4}$ に設定した。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。最終的な W と H が、 V をNBMFで分解して得られる解となる。図1のフローチャートは、NBMFのアルゴリズム全体を表したものである。

W は実数行列であるため、式(2)は射影勾配法 [20] を用いて更新する。 W の初期値は、構成する全ての要素が x_{\max} 以下であるような、任意の実数行列とする。本研究では、 $x_{\max} = 1.0$ に設定した。更新は行ベクトルごとに行う。求める W の行ベクトルを \mathbf{x}^T 、対応する V の行ベクトルを \mathbf{v}^T とおいて、次式を最小化する。

$$f(\mathbf{x}) = \|\mathbf{v} - H^T \mathbf{x}\|^2 + \alpha \|\mathbf{x}\|^2 \quad (5)$$

このとき、射影勾配法の更新式は以下の式になる。

$$\mathbf{x}^{k+1} = P[\mathbf{x}^k - \gamma_k \nabla f(\mathbf{x}^k)]$$

$$P[x_i] = \begin{cases} 0 & (x_i \leq 0) \\ x_i & (0 < x_i < x_{\max}) \\ x_{\max} & (x_{\max} \leq x_i) \end{cases} \quad (6)$$

ここで、 $\nabla f(\mathbf{x}^k)$ は $f(\mathbf{x})$ の勾配であり、

$$\nabla f(\mathbf{x}) = -H(\mathbf{v} - H^T \mathbf{x}) + \alpha \mathbf{x} \quad (7)$$

で求めることができる。はじめに $\gamma_0 = 1$ に設定して、

$$f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k) \leq \sigma \nabla f(\mathbf{x}^k)^T (\mathbf{x}^{k+1} - \mathbf{x}^k) \quad (8)$$

を条件として式(6)を計算する。定数 σ は、先行研究に倣い $\sigma = 0.01$ に設定する。 γ_k が条件(8)を満たす場合は、満たさなくなるまで γ_k / β を繰り返し、満たさない場合は満たすまで $\gamma_k \times \beta$ を繰り返す。定数 β は、先行研究に倣い $\beta = 0.1$ に設定する。更新が終わったら、新しい γ_k の値で式(6)を再度計算するという処理を繰り返し、最終的に、

$$\|\mathbf{x}^{k+1} - \mathbf{x}^k\| \ll 1 \quad (9)$$

を満たしたら、 \mathbf{x}^k を W の新しい行ベクトルとして設定する。この更新を全ての行で行うことによって、 W 全体を更新する。

行列 H は二値行列であるため、イジングマシンで更新式を解くことができる。式 (3) の解となる変数の数は $k \times m$ だが、イジングマシンで扱うことのできる変数には限りがある。そのため、 H の更新を列ごとに行うことによって、マシンに渡す変数を k まで減らす。

$$H_l = \arg \min_{\mathbf{q} \in \{0,1\}^k} \|V_l - W\mathbf{q}\|_2 \quad (10)$$

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 (10) を式 (11) の QUBO 表現に変換する。

$$f(\mathbf{q}) = \sum_i a_i q_i + \sum_{i < j} b_{ij} q_i q_j \quad (11)$$

$$a_i = \sum_r W_{ri}(W_{ri} - 2V_{rl}), \quad (12)$$

$$b_{ij} = 2 \sum_r W_{ri} W_{rj}. \quad (13)$$

このとき、QUBO の係数 (12)(13) を、表 1 のような行列形式でイジングマシンに渡す。

表 1 QUBO の係数から作成する行列

Table 1 A matrix made from QUBO coefficients.

	q_1	q_2	q_3	q_4	\dots	q_k
q_1	a_1	b_{12}	b_{13}	b_{14}	\dots	b_{1k}
q_2	0	a_2	b_{23}	b_{24}	\dots	b_{2k}
q_3	0	0	a_3	b_{34}	\dots	b_{3k}
q_4	0	0	0	a_4	\dots	b_{4k}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
q_k	0	0	0	0	\dots	a_k

すると、評価関数 (11) 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 (10) の計算を m 回行うことによって、 H の全体を更新する。

3. 使用する画像データについて

本論文における画像解析の実験では、The Olivetti faces dataset (AT & T Laboratories Cambridge) から取得した人間の顔画像のデータセットを利用した。

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、その番号を本章では「ラベル」と呼ぶ。各名につき 6 枚を、顔画像の特徴を取得するための訓練画像とし、1 枚を画像分類用のテスト画像とす

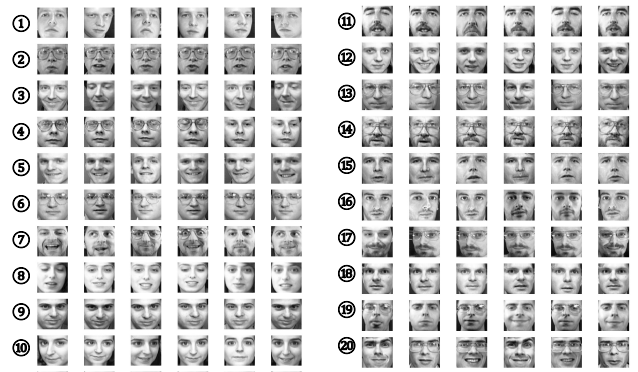


図 2 訓練画像 I_{train}

Fig. 2 Training image data.

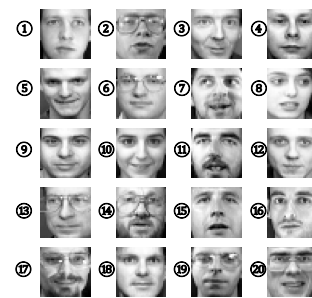


図 3 テスト画像 I_{test}

Fig. 3 Test image data.

る。従って、合計すると訓練画像が 120 枚、テスト画像が 20 枚となる。

本論文では、これらをまとめて画像データセット I と呼称し、訓練画像を I_{train} (図 2)、テスト画像を I_{test} (図 3) とする。

4. NBMF を用いる画像分類

4.1 画像行列の分解

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。今回の実験で利用した画像のサイズは、それぞれ 32×32 であるため、 V の行数は $n = 1024$ で、 V を構成する訓練画像は 120 枚であるため、列数 $m = 120$ となる。 $V \approx WH$ に近似するが、このとき、行列 W の列数かつ行列 H の行数にあたる特徴数 k は、任意の値に設定する。今回は $k = 60$ に設定した。根拠については後述する。

W の各列ベクトルが特徴にあたり、それぞれを画像として出力すると図 4 のようなものが得られる。顔画像から抽出した特徴が表れており、これらを重ね合わせることで元画像になるべく近いものを再現する。 H は、この再現のためにどの特徴を選んだかをバイナリの値で表す。



図 4 特徴行列 W を出力した画像

Fig. 4 A matrix W plotted as an image.

4.2 分類の手法

行列 V を分解して得られた W, H を利用する. I_{test} を構成する 20 枚の画像を, 行列 V_{test} に変換し, W で分解することによって H_{test} が得られる.

$$V_{\text{test}} \approx WH_{\text{test}} \quad (14)$$

H_{test} は, W を構成する顔画像の特徴の中から, どれを選んで V_{test} の再現を試みたかを表している. 従って, H の中に H_{test} の各ベクトルと構成が似ているベクトルが存在する場合は, 同じ様な特徴を選んでいるため, 元画像同士が似ていると考えることができる.

この性質を利用して, I_{test} の画像のラベルを推定する. その際, K 近傍法という手法を用いる. これは, ベクトル間の距離を利用してラベルの判定を行う手法である. H の i 番目の列ベクトルを $H^{(i)}$ とするとき, 行列を構成する全ての $H^{(i)}$ において, H_{test} 内におけるラベルを推定したい目的の列ベクトル \mathbf{h} との間の距離を, 次式の関数 distance で計算する.

$$\text{distance}(i) = \frac{\|\mathbf{h} - H^{(i)}\|}{k} \quad (15)$$

その中から, ベクトル間距離が短い順に, H から K 本のベクトルを抜き出す. それらに対応する元画像のラベルを参照し, 最も多く選ばれているラベルが, \mathbf{h} のラベルであると推定する. 例えば, $K = 3$ に設定すると 3 本のベクトルが H から選ばれるが, その内 2 本のベクトルに対応するラベルの番号が 1 で, 残りの 1 本は 2 番のラベルの画像であったとき, 多数決で \mathbf{h} は 1 番の画像であるとする. この判別を H_{test} を構成する全てのベクトルに対して行う.

今回は, I_{train} を同一条件で 5 回分解し, それぞれの分解で得られた訓練画像を用いて I_{test} を分類した. 5 回の分類における平均正解率は 93% となった.

5. Neural Network に基づく NBMF

5.1 画像行列の分解

NBMF に Neural Network の構造を導入する. 従来知られている Neural Network による分類手法では, 画像データを入力層として与え, 最終的にラベルの判別を示す出力層へと繋げる. その間に中間層 (隠れ層) が存在する. この構造を NBMF の行列表現で置き換えると, 大まかには

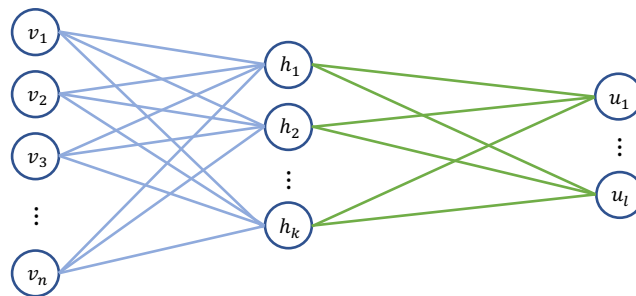


図 5 Neural Network の構造の概要

Fig. 5 The structure of Neural Network.

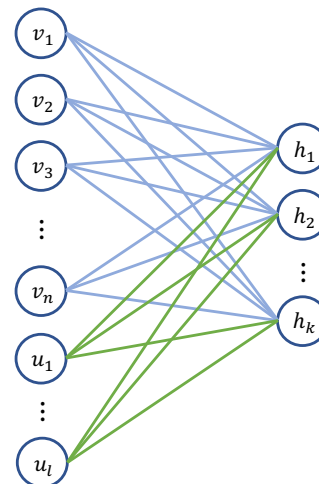


図 6 Neural Network を NBMF に適用した構造の概要

Fig. 6 The structure of NBMF based on Neural Network.

図 5 のような形になる. 訓練画像の行列 V の各列ベクトル \mathbf{v} を入力層とし, 中間層 \mathbf{h} を経て出力層 \mathbf{u} を得ることを目指す. \mathbf{u} はラベルの種類数と同じ $l = 20$ 次元のベクトルであり, 訓練画像のラベルを表現する. 対応する \mathbf{v} の元画像のラベル番号と同じインデックスの要素にだけ, 値 N を入れる. その他の要素はゼロで埋めることによってラベルの区別を行う. \mathbf{v} は画素値を規格化したベクトルであるため, 要素の値は最大で 1 である. N をそれより大きな数に設定することによって, この値がラベルを表すという差別化を図る. 今回の実験では, N の値は 5 に設定する. その根拠については後述する.

この構造を NBMF の行列分解のアルゴリズムに適応させるため, 図 6 のように, 出力層は入力層の下に結合する. 出力層を追加することによって, V の行数は $n + l = 1024 + 20 = 1044$ となる. これを分解して得られる行列 W のうち, 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. 従って W を, 特徴の部分である $n \times k$ 行列 W_1 と, ラベルの部分である $l \times k$ 行列 W_2 に分割する.

5.2 分類の手法

特徴行列を利用して, テスト画像の行列 V_{test} のラベル

判別を行う。 V_{test} を W_1 で分解すると、行列 H_{test} が得られる。

$$V_{\text{test}} \approx W_1 H_{\text{test}} \quad (16)$$

H_{test} はテスト画像の中間層にあたり、これに W_2 をかけることによって判別層 U_{test} を得る。

$$W_2 H_{\text{test}} \approx U_{\text{test}} \quad (17)$$

U_{test} の各ベクトルを、 V_{test} の元画像のラベル判別に利用する。 訓練画像の出力層では、対応する位置にだけ要素を入れることによってラベルを表したことから、各ベクトルの中で最も値の大きい要素のインデックスが画像のラベルであると推定する。 この判別を、 U_{test} を構成する全てのベクトルに対して行う。

先ほどの K 近傍法と同じく、5 回の画像分類を行ったところ、全ての実験で正解率が 100% となり、K 近傍法を利用するよりも優れた結果が得られた。

6. パラメータ設定

6.1 ビット数 k

NBMF での行列 W の列数かつ行列 H の行数となる k は、画像から抽出する特徴の数である。 k を大きな値に設定すると多くの特徴を得ることができ、詳細な画像解析を行うことができると考えられる。しかし、 k はイジングマシンに与える変数の数であるビット数でもある。マシンで扱えるビット数の数は限られている上に、大きすぎる値に設定すると一度に多くの変数を与えることになるため、精度が下がったり、計算に時間がかかる場合がある。従って k の値は、正確な画像解析は可能であるが、大きすぎない程度の数に設定する必要がある。また、NBMF は行列分解のアルゴリズムであるため、解として得られる行列の次元をもとよりも削減しているような k の値を設定したい。即ち行列 V だと、 $n \times m > k(n+m) \Leftrightarrow k < \frac{n \times m}{n+m}$ の条件を満たす値に設定する。今回の実験で分解する V の行数は $n = 1024$ 、列数は $m = 120$ であるため、代入すると $k < 107.4$ の整数となる。従って、 $10 \leq k \leq 100$ の範囲で 10 刻みで変化させ、パラメータの値以外は全て同一条件の画像分解を行う。分解前の画像行列 V と、分解の結果として得られる W, H の積との間で二乗平均平方根誤差 (Root Mean Squared Error; RMSE) を計算する。誤差が小さければ、分解前の画像に近いものを再現できているということになるので、この値を利用して最適化の精度を検証する。RMSE の推移をグラフ上にプロットしたものが図 7 である。

k を大きな値にしてたくさんの特徴を抽出するほど、画像の再現度が上がり RMSE が小さくなることがわかる。しかし、先述したビット数などの関係から、本研究では V の列数 m の 5 割の値に相当する $k = 60$ に設定する。グラ

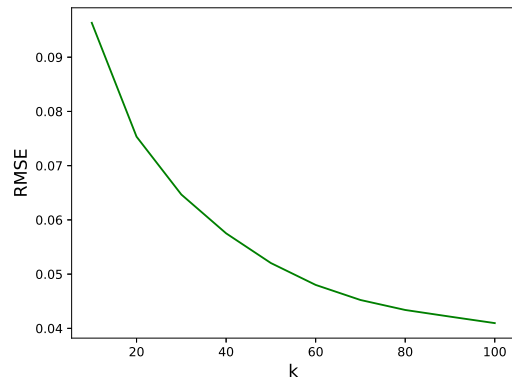


図 7 k を段階的に変化させた際の RMSE の推移
Fig. 7 RMSE plotted as a function of k .

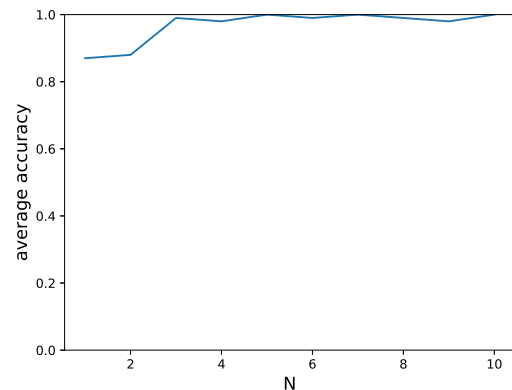


図 8 N を段階的に変化させた際の平均正解率の推移
Fig. 8 Average accuracy plotted as a function of N .

フより、 $k = 60$ 付近の RMSE は十分に小さく、また、以降の RMSE の減少も緩やかであることがわかり、適した値だと考えられる。

6.2 ラベル区別のための N の値

第 5 章で述べた Neural Network に基づく NBMF のアルゴリズムでは、出力層ベクトル u で画像のラベルを表す。 u の次元数は、データセット I のラベルの種類と同数であり、対応する入力層 v のラベル番号と同じインデックスの要素に N を代入する。 v の最大値は 1 であるため、 N はそれより大きい値にすれば判別可能であると考えられるが、最適な値を検証した。 $1 \leq N \leq 10$ の範囲の自然数で変化させ、各値で同一条件の行列分解による学習と分類を 5 回行い、平均正解率を算出した。その結果をプロットしたものが図 8 である。

グラフより、 $N \geq 3$ に設定すれば、安定して高い正解率が得られることがわかった。従って、今回は $N = 5$ に設定した。

7. おわりに

行列分解のアルゴリズムである NBMF を用いて、画像解析を行った。2 種類の分類手法を導入し、同一の画像

データセット I の特徴学習とラベル判別を行った。より良い結果が得られたのは Neural Network の概念に基づく手法であった。もう一方の K 近傍法による分類では、学習後に得られる特徴ベクトル間の距離を利用するのに対して、学習前の画像行列にラベル情報を加えてから行列分解を行う点が大きな違いである。ラベルの情報も踏まえた上で、NBMF を利用する分解によって特徴を抽出する方が、精度の良い結果が得られることがわかった。また、K 近傍法で判定を行うためには、テスト画像と全訓練画像との間で距離を計算し比較する必要がある。訓練画像数を増加させると解析の正確性は上がるが、ベクトル計算に要する時間も増えることになる。しかし、Neural Network に基づく手法では、テスト画像から得られる判別層のベクトルを参照するだけでラベルを推定することができる。計算回数を削減できるという点でも利便性がある。

謝辞 本論文の成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務 (JPNP16007) の結果得られたものである。

参考文献

- [1] Gibney, E.: D-wave upgrade: How scientists are using the world's most controversial quantum computer, *Nature News*, Vol. 541, No. 7638, p. 447 (2017).
- [2] Kadowaki, T. and Nishimori, H.: Quantum annealing in the transverse Ising model, *Physical Review E*, Vol. 58, No. 5, p. 5355 (1998).
- [3] Johnson, M. W., Amin, M. H., Gildert, S., Lanting, T., Hamze, F., Dickson, N., Harris, R., Berkley, A. J., Johansson, J., Bunyk, P. et al.: Quantum annealing with manufactured spins, *Nature*, Vol. 473, No. 7346, pp. 194–198 (2011).
- [4] Han, K.-H. and Kim, J.-H.: Genetic quantum algorithm and its application to combinatorial optimization problem, *Proceedings of the 2000 congress on evolutionary computation. CEC00 (Cat. No. 00TH8512)*, Vol. 2, IEEE, pp. 1354–1360 (2000).
- [5] Goto, H., Tatsumura, K. and Dixon, A. R.: Combinatorial optimization by simulating adiabatic bifurcations in nonlinear Hamiltonian systems, *Science Advances*, Vol. 5, No. 4, p. eaav2372 (2019).
- [6] Kong, D., Fan, X., Shmakov, K. and Yang, J.: A Combinational Optimization Approach for Advertising Budget Allocation, pp. 53–54 (2018).
- [7] Wang, L. and Zheng, D.-Z.: An effective hybrid optimization strategy for job-shop scheduling problems, *Computers & Operations Research*, Vol. 28, No. 6, pp. 585–596 (2001).
- [8] Inagaki, T., Haribara, Y., Igarashi, K., Sonobe, T., Tamate, S., Honjo, T., Marandi, A., McMahon, P. L., Umeki, T., Enbutsu, K. et al.: A coherent Ising machine for 2000-node optimization problems, *Science*, Vol. 354, No. 6312, pp. 603–606 (2016).
- [9] Yamaoka, M., Yoshimura, C., Hayashi, M., Okuyama, T., Aoki, H. and Mizuno, H.: A 20k-spin Ising chip to solve combinatorial optimization problems with CMOS annealing, *IEEE Journal of Solid-State Circuits*, Vol. 51, No. 1, pp. 303–309 (2015).
- [10] Matsubara, S., Takatsu, M., Miyazawa, T., Shibasaki, T., Watanabe, Y., Takemoto, K. and Tamura, H.: Digital annealer for high-speed solving of combinatorial optimization problems and its applications, *2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC)*, IEEE, pp. 667–672 (2020).
- [11] Neven, H., Denchev, V. S., Rose, G. and Macready, W. G.: Qboost: Large scale classifier training with adiabatic quantum optimization, *Asian Conference on Machine Learning*, PMLR, pp. 333–348 (2012).
- [12] Benedetti, M., Realpe-Gómez, J., Biswas, R. and Perdomo-Ortiz, A.: Quantum-assisted learning of hardware-embedded probabilistic graphical models, *Physical Review X*, Vol. 7, No. 4, p. 041052 (2017).
- [13] Hu, F., Wang, B.-N., Wang, N. and Wang, C.: Quantum machine learning with D-wave quantum computer, *Quantum Engineering*, Vol. 1, No. 2, p. e12 (2019).
- [14] Neukart, F., Von Dollen, D., Seidel, C. and Compostella, G.: Quantum-enhanced reinforcement learning for finite-episode games with discrete state spaces, *Frontiers in Physics*, Vol. 5, p. 71 (2018).
- [15] Li, R. Y., Di Felice, R., Rohs, R. and Lidar, D. A.: Quantum annealing versus classical machine learning applied to a simplified computational biology problem, *npj Quantum Information*, Vol. 4, No. 1, pp. 1–10 (2018).
- [16] O'Malley, D., Vesselinov, V. V., Alexandrov, B. S. and Alexandrov, L. B.: Nonnegative/binary matrix factorization with a d-wave quantum annealer, *PLOS ONE*, Vol. 13, No. 12, p. e0206653 (2018).
- [17] Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P.: Optimization by simulated annealing, *Science*, Vol. 220, No. 4598, pp. 671–680 (1983).
- [18] Hukushima, K. and Nemoto, K.: Exchange Monte Carlo method and application to spin glass simulations, *Journal of the Physical Society of Japan*, Vol. 65, No. 6, pp. 1604–1608 (1996).
- [19] Asaoka, H. and Kudo, K.: Image Analysis Based on Nonnegative/Binary Matrix Factorization, *Journal of the Physical Society of Japan*, Vol. 89, No. 8, p. 085001 (2020).
- [20] Lin, C.-J.: Projected gradient methods for nonnegative matrix factorization, *Neural Computation*, Vol. 19, No. 10, pp. 2756–2779 (2007).