

2025 年度大学院入試 工学研究科 博士後期課程(後期日程) 出題意図

一般選抜【専門科目】＜機械システム工学領域＞

問[Ⅰ]は, 機械システム工学において重要とされる画像処理技術に関連して, パターン識別にかかわる基本的な知識を問うものである. まず(1)では線形代数に関わる基本知識として主成分分析, (2)ではさらにその応用として PCA の利用法に関する知識を問う. (3)および(4)では PCA の発展的手法でもある部分空間法の考え方および知識を問う, 最後の(5)では特徴空間における距離計算等でよく用いられるマハラノビス距離に関する基本的な知識を問う.

問[Ⅱ]では, 数学の基本として, 線形写像に関する知識を問う(1), 関連の計算ができるかどうかを問う(2). また(3)ではクラス分類技術の位置づけや実応用に関する知識を問う.

問[Ⅲ]では, ロボットビジョン(視覚)で頻繁に利用されている遺伝的アルゴリズム(GA)の知識を問うものである. (1)では GA の基本的な仕組み, (2)では実際によく利用される交叉に関するごく基礎的な知見を問う. (3)では GA のポジショニングに関する知識を問う, (4)では実問題に GA を適用する際に必要な知識を問う. さらに(5)では, GA の利用場面を具体的に想定できているかを問う.

解答例

2025 年度大学院入試 工学研究科 博士後期課程（後期日程）

一般選抜【専門科目】＜機械システム工学領域＞

〔I〕パターン識別（Pattern Classification）における部分空間（Sub-space）法に関連して、
（1）～（5）の各問いに答えよ。ただし、次のすべてのキーワードを（1）～（5）の解答のいずれか
の中に含めること（出現順序は問わない）。必要に応じて数式や図を用いてもよい。
文字数は問わないが、日本語で解答すること。

【使用すべきキーワード】

固有方程式，第1主成分，線形変換，累積寄与率，分散共分散行列，正規直交基底

- （1）主成分分析（PCA）について説明せよ。
- （2）PCAによる次元削減における削減後の適切な次元数の決定方法について論ぜよ。
- （3）SELFIC法について説明せよ。
- （4）CLAFIC法について説明し、SELFIC法との使い分けの方針について論ぜよ。
- （5）マハラノビス距離について説明せよ。

- （1）主成分分析（PCA）は、多次元空間内のデータを低次元空間に写像するための線形変換手法の一種である。おおまかな手順としては、元のデータから分散共分散行列を計算し、これを分解して固有値と固有ベクトルを求める。このとき、固有方程式を解くことになる。固有ベクトルのうち、最も大きい固有値に対応する固有ベクトルを第1主成分といい、以下、第2主成分、第3主成分・・・と続く。得られる固有空間は元の空間より次元が低く、結果としてデータ量の削減が可能である。なお、主成分分析によって得られる空間は、正規直交基底からできている。
- （2）PCAにおける各固有値はそのベクトル（次元）の情報としての重みを表現している。固有値を大きい順に並べ、各固有値の累積寄与率を調べることによって、上位何次元までを利用すればどのくらい元の情報を維持できるかを知ることができる。これに基づき、たとえば、累積寄与率が98%になるときの次元数 M を調べることによって、適切な次元数は M である等の決定がなされる。一般に累積寄与率グラフは、早い段階で急峻に立ち上がり、累積寄与率は徐々に100%に漸近するような形を持つことが多い。これはごく少ない次元数で高い寄与率を達成可能なことを示している。
- （3）SELFIC法は、画像等のデータをクラス分類するための線型識別手法の一種である。複数のクラスにわたる教師付きデータを主成分分析によって低次元化し、その空間（部分空間）において、未知のテストデータと、既知のクラスごとのプロトタイプデータとの距離等を計算することによって識別をおこなう。
- （4）CLAFIC法もまた、画像等のデータをクラス分類するための線型識別手法の一種であるが、こちらは複数のクラスごとに主成分分析をおこない、クラスごとに部分空間を作る。その上で、未知のテストデータを各部分空間に射影したベクトルの大きさをもとに、これが大きい場合にはテストデータがそのクラスに属する可能性が高いと判断する。一般には、学習用の各クラスのデータ分散が大きい、またはクラスごとのデータの分散値が異なる場合（例えば顔認識）にはCLAFIC法を用いることが多い。

- (5) マハラノビス距離は，多次元空間におけるデータ間の距離測定手法の一種であり，いわばデータ間の類似性を算出するための指標である．多次元空間においては，しばしば軸ごとにデータの分散値が大きく異なり，これに起因してユークリッドノルム等の他の距離測定指標が不適な場合がある．マハラノビス距離は軸ごとのデータの分散値を考慮して分散値の差異を緩和するため，このような場合にも適した指標となっている．

〔Ⅱ〕写像について，(1)～(3)の各問いに答えよ．日本語で解答すること．

(1) ある写像が線形写像 (Linear Projection) であるための条件を答えよ．

(2) $\mathbf{R}^4$ から $\mathbf{R}^3$ への写像 (Projection)

$$F: (x_1, x_2, x_3, x_4) \mapsto (x_2, x_1, x_4)$$

は線形写像であるかどうかを判定せよ．

(3)線形写像の AI (人工知能) への応用について，知るところを述べよ．

- (1) ある写像 F が線形写像であるための条件は，任意の x, y, λ について，

$$F(x + y) = F(x) + F(y)$$

$$\lambda F(x) = F(\lambda x)$$

を満たすことである．

- (2) 空間 $\mathbf{R}^4$ における写像前の 2 つの座標 (ベクトル) を $x = (x_1, x_2, x_3, x_4)$, $y = (y_1, y_2, y_3, y_4)$ とする．

$$F(x + y) = F(x_1 + y_1, x_2 + y_2, x_3 + y_3, x_4 + y_4) = (x_2 + y_2, x_1 + y_1, x_4 + y_4)$$

$$F(x) + F(y) = (x_2, x_1, x_4) + (y_2, y_1, y_4) = (x_2 + y_2, x_1 + y_1, x_4 + y_4)$$

$$\lambda F(x) = \lambda(x_2, x_1, x_4) = (\lambda x_2, \lambda x_1, \lambda x_4)$$

$$F(\lambda x) = (\lambda x_2, \lambda x_1, \lambda x_4)$$

以上より， F は線形写像である

- (3) AI 分野の代表的技術に，物体識別，クラス分類と呼ばれる技術がある．これにはさまざまな手法があるが，教師有り学習に基づく手法の例として，あらかじめ用意した教師信号付きの学習用画像群を，画像空間，あるいはなんらかの方法で特徴量化した特徴空間で表現し，これを線形写像により低次元化した新しい空間を生成した上で識別するという方法がある．線形写像は計算が容易であり，低次元化により処理の高速化や，場合によっては精度向上も狙えることから，重要な基本技術となっている．線形写像は行列の乗算により実行されるので，GPU (Graphic Processing Unit) のような CPU 外のアクセラレータチップによって高速演算することもでき，有用性が高い．

〔Ⅲ〕組合せ最適化 (Combinatorial Optimization) 法の一つである遺伝的アルゴリズム (GA : Genetic Algorithm) に関連して，(1)～(5)の各問いに答えよ．ただし，次のキーワードをすべて使用すること (出現順序は問わない)．必要に応じて数式や図を用いてもよい．文字数は問わないが，日本語で解答すること．

【使用すべきキーワード】

個体，染色体，交叉，突然変異，致死遺伝子，適応度，世代交代，収束

- (1) 遺伝的アルゴリズム (GA) の基本的な仕組みについて説明せよ．

(2) 一様交叉について説明せよ.

(3) GA 以外の組合せ最適化法を 1 つ挙げて原理を説明し, GA との長所短所を比較せよ.

(4) 交叉率および突然変異率を随時変更する場合の考え方について論ぜよ.

(5) どのような組合せ最適化問題に GA の利用が適しているかについて論ぜよ.

- (1) 遺伝的アルゴリズムは, 組合せ最適化問題に対する近似解法の一つであり, 生物の進化過程を模していることが特徴である. まず最初に, 解を染色体にコーディングしたうえで, ランダムな染色体をもつ多数の初期個体群を生成し, 最適化の目的関数に相当する適応度 (fitness) を計算する. 適応度の高い 2 つの親個体をもつ 2 つの染色体同士を遺伝的操作によって掛け合わせる. 遺伝的操作としては, 一般に, 交叉と突然変異が用いられる. その結果, 新たな子個体が生成され, 一定の基準に沿わないものは致死遺伝子として消去されるが, そうでないものは適応度が計算され, 個体群に戻される. このとき個体群の中で最も適応度が低い個体を消滅させることによって, 個体群の総数を維持するのが一般的である. このようにして得られた子個体をもつ染色体は, 優良な親個体の性質を受け継いでいることが期待されることから, しばしば親個体より高い適応度を持つ. このような世代交代を繰り返すことによって個体群が収束し, 最終的には, きわめて適応度の高い個体, すなわち高い目的関数を有する解が出現することを期待するアルゴリズムとなっている.
- (2) 一様交叉 (uniform crossover) は, GA における交叉の一種であり, 染色体のビット列と同じ長さの 2 値のランダムビット列を生成し, 値 1 を持つビットは親 A のビットが, 値 0 をもつビットは親 B のビットがコピーされて子個体が生成される. 通常, 子個体は 2 つ生成されるので, もう一方については, 前記と逆の親のビットがコピーされる. 交叉の手法には, 染色体中の特定の 1 つのビットの前後によってどちらの親個体のビットをコピーするかを決める 1 点交叉や, これを 2 つのビットに拡張した 2 点交叉などもあるが, 解の染色体へのコーディング手法によっては, ビットの位置そのものに意味を持たせない場合も多いことから, その場合にはこの一様交叉が有効である.
- (3) 組合せ最適化法には, 焼き鈍し法 (SA : Simulated Annealing), 分枝限定法 (Branch and Bound) などさまざまな手法がある. たとえば SA 法は greedy な手法であるとともに, 解を一定の確率で変化させる確率的要素を含んでいる. この点においては, GA もまた, 交叉のような greedy な効果と, 突然変異のような確率的な効果の両方を狙っていることから, 類似していると言える. SA 法と比較した GA の利点としては, アルゴリズムそのものが単純な考え方で作られていることから, コンピュータプログラムとして実装しやすい, 目的関数を必ずしも数値関数として表現する必要はなく, 適応度が定義できればよいことが挙げられる. 一方では, 解の染色体へのコーディングが不適切な場合には, 交叉等の遺伝的操作が「よりよい解」を生む結果につながらないことがある, 交叉率や突然変異率のようなパラメータ調整に手間がかかる場合があるという問題点も考えられる.
- (4) 遺伝的操作のうち, 交叉は, 準最適解を部分的に組み合わせてよりよい解を得ることが期待される処理であるから, 世代交代の初期段階においては, 交叉の頻度は高めに設定されるべきである. 初期段階における個体群には, かなり多様性の高いさまざまな個体 (染色体) が存在していると考えられることから, この組合せの効果が大きい. 一方, 突然変異については, 染色体ビットのなかで, 1 つまたは複数のビットを確率的に変更することによって, greedy な求解では得られなかった新規性の高い解を生成することを期待していることから, 初期個体群の多様性によらず効果が期待できる. このことから突然変異は世代交代の後半で行われるほうが効果が高いと考えられる. 実際の GA の運用にあたっては, たとえば個体群中のすべての染色体がもつ適応度の分布をモニタリングし, 平均適応度と最大適応度に十分な差がある状況では交叉率を大きく, 突然変異は小さくし, 世代交代が進み, 平均適応度と最大適応度の差が小さくな

ってきたタイミングで、個体群の多様性が低減したと判断し、徐々に交叉率を小さく、突然変異率を大きくするのがよいと考える。

- (5) GA の大きな特徴の一つは、準最適解の組合せによってよりよい解が得られる可能性が高いであろうという点にある。これは必ずしも保証されたことではないが、一般に **Building-Block** 仮説と呼ばれ、十分に期待できる仮説である。このことから、解きたい問題を、染色体、具体的には 1 次元のビット列からなるストリングスに適切にコーディングでき、この仮説を満たしやすい場合に適していると考えられる。また、組合せ最適化手法の中には、目的関数を数値関数で表現することを必要としたり、場合によっては連続関数であることを要求しているものもあるが、GA については、目的関数の値は適応度という数値指標で表現できさえすればよく、必ずしも数式で計算される必要はない。すなわち、コンピュータプログラムの中で、適応度を数値として出力できるモジュールを構成できればよい。言い換えれば、目的関数を数式で表現しにくい場合には、GA の利用が適している。