

情報学専攻入学者選抜試験

行動情報学

問題冊子

注意事項

- (1) 試験開始の合図があるまで、この問題冊子の中を見てはいけません。
- (2) 問題冊子は、本表紙・空白を含め、全部で 12 ページあります。試験開始後に問題冊子のページ数を確認して、落丁又は印刷が不鮮明な場合は直ちに申し出てください。
- (3) 試験開始後に、問題冊子・解答冊子の表紙と解答用紙のすべてのページに受験番号を記入してください。
- (4) 問題冊子、解答冊子は切り離さずにすべて提出してください。
- (5) 解答は解答冊子に記入してください。問題冊子に記入しても採点されません。
- (6) 「情報処理」分野はすべての問題に解答してください。「情報システムとデータ分析」分野は大問 3 問から 2 問を選んで解答してください。
 - 選択解答した大問がわかるように、**解答冊子表紙**に記載してある「情報システムとデータ分析」分野の大問番号欄の□内に印をつけてください。
 - 「情報システムとデータ分析」分野の 3 問すべてに印をつけた場合は、すべて無効になることがあります。

分 野	解 答	備 考
情報処理	すべての問題を解答	
情報システムとデータ分析	大問 3 問から 2 問を選択して解答	解答冊子表紙の解答した大問 番号欄の□内に印をつける

受験番号	
------	--

情報処理

※全問を解答すること。

（問１）

相手にファイルを送る際、次の手順で実施することがある。ここで、送信対象のファイルを **file** とする。

- ① 送り主は、**file** を含むパスワード付き **ZIP** ファイルを作成し、電子メールで相手に送信する。
- ② 送り主は、①で用いたパスワードを、①とは別便の電子メールで相手に送信する。
- ③ 相手は、①で受け取った **ZIP** ファイルを②で受け取ったパスワードを使って展開し、**file** を入手する。

この方法は **PPAP** と呼ばれ広く用いられるようになってきたが、幾つかの観点から危険性が指摘されている。**PPAP** について、次の各問いに答えよ。

- （１）**PPAP** の危険性について説明せよ。
- （２）**PPAP** の危険性を回避するためにどのような方法が考えられるか説明せよ。

（問２）

インターネットに接続されたコンピュータ類（ホスト）を識別するために用いられる **IP** アドレスについて、グローバル **IP** アドレスとプライベート **IP** アドレスがある。グローバル **IP** アドレスに加えてプライベート **IP** アドレスを利用することの利点について説明せよ。

（問３）

次の **I** ～ **IV** の各記述に該当する用語を解答群の中から一つ選べ。

- (I) **CPU** と主記憶装置の間に高速な記憶装置を配置して、主記憶の実効アクセス速度を向上させる。
- (II) オペレーティングシステムが記憶領域の割当てと解放を繰り返すと、こま切れの未使用領域が多数できる。
- (III) 主記憶を複数の領域に分割し、複数の領域に並列にアクセスすることによって、主記憶へのアクセスを高速化する。
- (IV) ページング方式の仮想記憶システムにおいて、ページ置換えの発生頻度が高くなると、システムの処理能力が急激に低下する。

解答群

キャッシュメモリ
セグメンテーション
ページング
ラウンドロビン

スラッシング
フラグメンテーション
メモリアンターリーブ

(問4)

以下の文を読んで、設問に解答せよ。

Tさんは、Web ページの内容を解析してそこから他の Web ページへのリンク先 URL を取り出して集計するプログラムを作っている。Tさんの作っているプログラムでは、与えられた複数の Web ページのソースを単一のテキストファイルになるように結合したあと、その結合したテキストファイル中に出現する URL を文字列 (String) 型のオブジェクトとして先頭から順に取り出し、別途用意した文字列型の配列に対してそれを出現順に入れるようにしている。

- (1) ここで配列というデータ構造を用いることによって有利になる点や不利になる点には何があるか。そのかわりに連結リストなどの他のデータ構造を用いた場合と比較して述べよ。
- (2) このプログラムを実際に動作させ、そこで配列に格納した URL の数が n 個であるとき、別途指定された URL とは一致しないものをこの配列内からすべて調べて表示するような具体的なアルゴリズムを、擬似コードまたは Java コードなどで示せ。ただし、ここで指定された URL とは文字列型の変数 `target` として与えられているものとし、プログラム中で URL を格納する文字列型の配列 `stored[]` には、0 番目から $(n - 1)$ 番目の要素にそれぞれ URL を示す文字列がその出現順に格納されているものとする。
- (3) 上述の (2) の操作を行う具体的なアルゴリズムにおける、その処理にかかる計算量を n に対するオーダー $O(\dots)$ の形で示せ。ただし、文字列の比較を行う操作の時間計算量は $O(1)$ としてよいこととし、 n は格納した URL の個数であるとする。
- (4) URL を格納する時に、すでに格納した URL は重複して入れないようにして、そのかわりにその URL の出現回数を数えるようにしたい。そのためには、このデータ構造を文字列型の配列からどのようなものに変更すればよいかを示せ。
- (5) 上述の (4) の変更をした際に、すでに格納された URL の種類数が m であるとき、指定された URL がすでに出現したものであるかを、すでに格納された URL の集合に対して調べるための時間計算量がどのようになるか、その理由とともに示せ。

情報システムとデータ分析

※問1～問3から2問を選択して解答すること。解答冊子の表紙の選択した問題の□内に印をつけること。

（問1）

A社は、地方都市で、観光客向け商品を中心に和菓子を製造している従業員80人の企業である。その強みは、その都市で長年親しまれてきた商品のブランド力である。一方、販売に関しては、直営店に加え、卸売業者経由で鉄道の駅や高速道路のサービスエリアでの販売を中心に行ってきた。今後、ブランドの強みを生かして拡販を行うため、洋菓子にアレンジした商品も含め、ECサイトでのオンライン販売専用商品を6種類追加し、日本全国に販売を行うことを計画している。従来の販売ルートでの売り上げを維持しつつ、新たなオンライン販売により、売上増加を狙うものである。

以下の設問に解答しなさい。

- （1） 経営学者アンゾフは、成長戦略を「現在の市場／新しい市場」と「現有製品／新製品」の二軸からなる成長マトリクスを用いて説明している。A社のオンライン販売導入について、この成長マトリクスの考え方を用いて説明しなさい。
- （2） A社が新たに始めるオンライン販売事業では、従来の卸売業者に対する販売と異なり、不特定多数の顧客を対象としているため、ECサイトへの誘導を図ることを目指している。さらに、ECサイトから顧客自身が入力することにより、商品購入と代金決済が完結するオンライン販売システムを構築する。これにより、確実な入金管理が実現できるだけでなく、顧客の購買実績データをもとにしたマーケティング活動の展開につなげていくことを考えている。このためには新たにどのようなデータ項目が必要になるのかについて述べなさい。
- （3） A社がオンライン販売事業を行うためには、オンライン販売のシステムを構築する必要がある。A社は、システム会社であるB社に開発を委託し、C社のクラウドコンピューティング環境を用いてシステム開発プロジェクトを行うこととなり、6か月後にサービス開始することとした。A社は、B社及びC社に対して、どのようなプロジェクトマネジメントを実施していくことが望ましいのか、論点を3点挙げて説明しなさい。
- （4） ビジネスをモデル化する手法の一つとして Unified Modeling Language（以下UML）がある。A社のオンライン販売に関して、ECサイトに初めてアクセスした顧客を想定してUMLのユースケース図を描きなさい。

(問2)

教師付き機械学習とは、「正解」とみなせる入出力ペア集合の訓練データを用いて学習を行い、未知の入力データに対しても推測する出力を行うものである。機械学習の性能評価には、利用可能な入出力ペア集合のうち一部を訓練データとして用い、残りを評価データとすることが多い。すなわち、評価データにおける推測結果のどのくらいが「正解」と一致するかを測る。

例として、工場で生産するある製品が良品か不良品かの二値に分類する問題を考える。実際に機械学習を行い、結果を評価したところ、下記の4つの評価尺度についてその件数が、

TP (True Positive、真陽性) = 50

TN (True Negative、真陰性) = 20

FP (False Positive、偽陽性) = 20

FN (False Negative、偽陰性) = 10

であったとする。なお、良品を陽性、不良品を陰性として扱う。

- (1) 再現率 (recall)、適合率 (precision)、正解率 (accuracy) の3つの評価尺度を計算せよ。小数点以下4桁目を切り捨てること。
- (2) (1) で計算した3つの評価尺度の意味をそれぞれ簡潔に説明せよ。また、(1) の計算結果はどう解釈できるか、それぞれの評価尺度とその関係について簡潔に説明せよ。

この工場で生産される製品は、型に熱した材料を流しこんで成型するプラスチック製品である。

製品について良品か不良品かの正解付きデータセットを 100 件準備できたとする。それぞれのデータセットには以下のようなデータが含まれている。

- a. 製品製造を行う作業者が製品製造過程を記録したテキストメモ（製品サンプルごとに 1000 文字程度）
- b. それぞれの製品を製造した時点の気温と湿度を記録した測定数値
- c. 角度を変えつつ様々な方向からそれぞれの製品を撮影した 1 分程度の動画

用意したデータを使い、機械学習手法を用いてこの製品が良品か不良品かを自動推測するシステムを実装したい。

- (3) この自動推測システムを実装するための設計を行い、使用する学習手法、訓練・評価時のデータ設定、学習結果の評価尺度を、それぞれ採用した理由を添えて、実装方法にかかわらず同じ結果が得られる程度にできるだけ詳細に記述せよ。設計には必ず、交差検証を含めるようにせよ。記述した設計を実装するために必要な技術仕様（実装に用いるソフトウェアライブラリなど実行環境、必要なハードウェア要件等）も記述に含めること。必要に応じて図を用いてもよい。
- (4) (3) で実装したシステムを用いて良品か不良品の推測を行った結果、(1) で計算した評価結果を得たとする。推測性能を向上させるために有効と考えられる改善策を 3 つ以上列挙せよ。それぞれ簡単に理由を添えること。

(問3)

- (1) S 県のプロサッカーチームは、20XX 年シーズンに J2 リーグを戦っており、勝利を重ねて早く J1 リーグへ復帰することが渴望されていた。データに基づいてこのような問題の解決をおこなう場合、その課題を科学的なフレームに落とし込むために、客観的なデータ指標（勝利数等）を定め、その値の変化や変動に何が影響するのかを、論理的なモデルに記述して問題の全体像を把握する必要がある。

論理モデルを記述する代表的な方法の一つに特性要因図（フィッシュボーンチャート）がある。地球の平均気温について記述した特性要因図のサンプルを図 1 に示す。解答用紙に示す特性要因図において、必要と考えられるだけ四角の空欄を埋め、さらに矢印などを追記することによって特性要因図を完成させなさい。また、完成させた特性要因図を説明することで、試合の勝利数増加につながる要因を明らかにしなさい。

なお、サッカーに詳しくない場合には、他のチームスポーツ（野球、バレーボール、バスケットボール等）の勝利数に事例を変更してもかまわない。

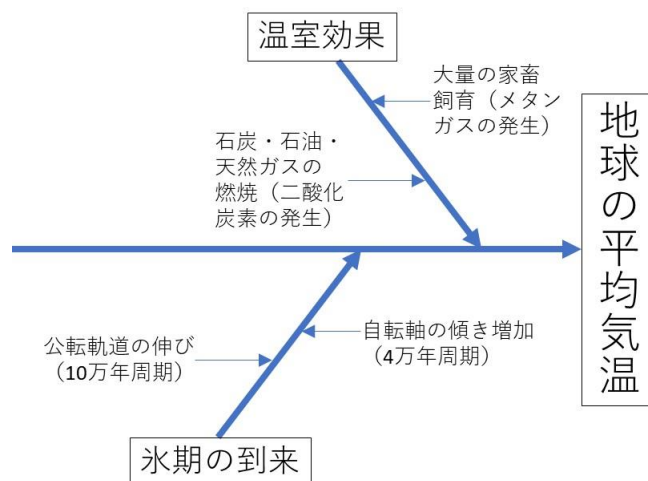


図 1. 特性要因図のサンプル（地球の平均気温）

- (2) ある大学の某研究室には、5 人の大学院生が所属している。彼らのアルバイト収入について調べたところ、表 1 のようになった。

4 月のアルバイト収入と 5 月のアルバイト収入の平均および標準偏差を求めなさい。また、4 月のアルバイト収入と 5 月のアルバイト収入について相関係数を求めなさい。

表 1. 大学院生のアルバイト収入調査結果

名前	4 月（千円）	5 月（千円）
A	70	85
B	80	75
C	55	70
D	50	65
E	50	55

- (3) 某企業では、カフェ 10 店舗をチェーン展開している。その営業実績データは表 2 のようになっている。このデータをもとに売上高に影響を与える要因を重回帰分析で調べたところ、表 3 のような結果を得ることができた。表内の「 < 0.001 」とは、値が 0.001 よりも小さいことを示している。

重回帰分析の結果について説明しなさい。その際、「自由度調整済み決定係数」、「 F 検定における p 値」、「回帰係数」、「 t 検定における p 値」の用語を用いること。

表 2. 某カフェチェーンの店舗別営業実績データ

店舗名	席数 (席)	最寄り駅からの 徒歩時間数 (分)	朝食セットの有無 (有：1，無：0)	売上高 (千円/年)
カフェ_1	78	8	1	84,340
カフェ_2	50	7	0	75,810
カフェ_3	70	2	1	92,880
カフェ_4	45	9	0	77,490
カフェ_5	76	1.5	1	93,960
カフェ_6	40	7	0	72,350
カフェ_7	85	4	1	85,960
カフェ_8	80	1.5	1	94,010
カフェ_9	75	3	1	87,180
カフェ_10	54	6	0	78,000

表 3. 重回帰分析結果

重相関係数	0.949
決定係数	0.900
自由度調整済み決定係数	0.850
標準誤差	3085.765
観測数	10

分散分析表

	自由度	変動	分散	観測された分散比	F 検定における p 値
回帰	3	515980689.434	171993563.145	18.063	0.002
残差	6	57131670.566	9521945.094		
合計	9	573112360.000			

	回帰係数	標準誤差	t 値	t 検定における p 値
切片	85442.291	10105.705	8.455	< 0.001
席数 (席)	-5.387	200.291	-0.027	0.979
最寄り駅からの 徒歩時間数 (分)	-1279.348	518.087	-2.469	0.048
朝食セットの有無 (有：1，無：0)	8960.439	6763.230	1.325	0.233

