

記述式小テストにおいて解答の傾向を
把握するための分類手法に関する研究

平成18年度

三重大学大学院工学研究科
博士前期課程 電気電子工学専攻

中川 敦志

修士論文

記述式小テストにおいて解答の傾向を
把握するための分類手法に関する研究



平成18年度修了

三重大学大学院工学研究科

博士前期課程 電気電子工学専攻

中川 敦志

目次

第1章	はじめに	1
第2章	記述式小テスト	3
2.1	小テスト	3
2.2	記述式小テストの解答の取り扱い	5
第3章	文書処理	7
3.1	文書の特徴ベクトル	7
3.1.1	tf・idf法	7
3.1.2	大井らの方法	9
3.2	文書分類における多変量解析	12
3.2.1	階層型クラスタリング	12
3.2.2	非階層型クラスタリング	14
3.2.3	低次元表現	15
第4章	提案法	19
4.1	記述式小テストの解答の傾向を把握するための分類	19
4.1.1	各手法による解答の分類	20
4.2	提案する分類法	22
4.3	クラスタの内容の表示	28
第5章	実験	30
5.1	実験条件	30
5.2	実験結果	32
5.2.1	考察	38
第6章	まとめ	40
	謝辞	41

参考文献	42
付録 A 実験で用いた解答	44

第1章 はじめに

講師がわかりやすい授業をするためには、学生が授業の内容をどの程度理解できているかを知ることが重要である。例えば、学生が授業の内容を理解できていないまま授業が進むと、それ以降の授業が理解できない場合があるからである。

これを解決するためには、講師は分からない部分があれば学生に質問させることや授業中に小テストを行うことが効果的である。学生に質問させることについては、学生が積極的に質問しない場合がある。そのため、講師が学生の理解の程度を把握できない場合がある。小テストを行うことについては、講師が小テストの解答を見ることで、学生がどの程度授業の内容を理解しているかがわかる。小テストには、選択式や穴埋め式、記述式などの小テストがある。選択式や穴埋め式の小テストでは、授業の内容を理解していなくても正解できる場合があり、学生の理解度を把握することが難しい。一方、記述式の小テストでは、学生自身の言葉で書かなければならず、授業の内容を十分に理解していないと正解できない。しかし、記述式小テストでは、数十人いる学生の解答すべてに目を通し、全体を把握するには時間がかかりすぎる問題がある。

そこで、本研究では講師が全学生分の記述式小テストの解答を即座に把握することを助けるために、解答を内容にしたがって解析を行い、解答の傾向を把握しやすいように解答にどのような内容があるかを提示する方法を提案する。このような文章で記述された情報を取り扱うことは、文書の特徴を表すように特徴ベクトルを作成し、特徴ベクトルに解析手法を用いることで実現できる。このような文書処理については多くの研究がなされている [1][2][3][4]。

講師が、小テストの傾向を把握するためには、講師は解答群の中から同じ内容の解答がいくつあるか、それが何種類あるかを知る必要がある。そのためには、解析した結果により、小テストの解答群の内容、その解答数の概数がわかればよい。

特徴ベクトル作成の技術では、既存の技術 [4][5] は、対象が数百万件の文書に用いるものであり、本研究の対象としている、数十件の文書とは件数が大きく違う。また、小テストの解答は、同じ質問の解答なので、内容のわずかな差に基づき分

類する必要がある。このことが、大きく影響するのはベクトルを作成する部分である。文書数が多いということは、使用される単語数も多くなり各文書のベクトルの情報量が大きくなる。このため、文献[5]のtf・idf法でも、各文書の違いが明確になりやすいが、小テストの解答は、数十件なので情報量が少なく、各ベクトルの違いが小さくなる。この問題を解決するために、文献[1]の大井らの方法では1件あたり数行程度、100件前後の文書から、分類の精度を高めるための情報を抽出するベクトル化の手法について検討している。本研究では、これを用い解答のベクトル化を行う。

特徴ベクトルの解析手法では、同じ内容の解答を1つにまとめ、その中からその内容を表す情報を表示し、同じ内容の解答群の数を表示することで解答数を知ることができる。しかし、小テストの解答が何種類存在するかという情報はない。そのため、解析結果の解答群の数が実際の解答の種類より多くなり解答数を把握できなくなる場合がある。この問題を避けるために、本研究では解析結果を視覚的に表示する方法を検討し、それを実現する方法を提案する。また、解析結果の解答群から内容を表す情報を表示することに関しては、解答群の内容を要約し表示することや、使用されている頻度が高い単語を表示するなどさまざまな方法が考えられる。本研究では、解析アルゴリズムの特性を使用して解答群の中から代表の解答を選ぶ方法を提案する。

結果については、小テストの解答の多数を占める内容の解答に対して、代表の解答が選ばれその概数がわかるようになった。

本論文の構成を以下に示す。2章では記述式小テストについて述べ、3章では文書の分類法について述べる。4章では本研究で提案する手法を示し、5章で実験によって提案法の評価をする。最後に6章で本論文をまとめる。

第2章 記述式小テスト

ここでは、講師が学生にとってわかりやすい授業をするために、小テストがなぜ効果的なのかを述べる。特に記述式小テストの重要性について述べる。また、講師が小テストの解答から、学生の理解の程度を知り授業の内容を改善するために必要な情報について検討する。

2.1 小テスト

講師が学生にとってわかりやすい授業をするためには、学生が授業の内容を理解しているかを知ることが重要である。なぜなら、学生が授業の内容を理解していなければ、それから先の授業も理解できなくなる場合があるからである。これを防ぐためには、講師が学生の理解の程度に基づいて追加説明などのフィードバックをして授業を行うことが必要である。そして、このフィードバックをできるだけ早く行うことが望ましい。

講師が学生の授業の理解の程度を知るためには、

1. 学生に質問をさせる。
2. 小テストを実施する。

などを行わなければならない。ここで、小テストとは講師が授業中に行うテストのことで、成績に関係なく学生の理解の程度を把握するためのものであり正誤は見ないものとする。1. では学生は積極的に質問を述べない場合や、学生が授業の内容を間違っ理解していても、自分で理解していると判断し質問しない場合があるので適さない。逆に、2. では、学生の積極性に関係なく、学生は授業の内容を理解していないと解答できない。以上より、講師が学生の授業の理解の程度を知るためには小テストを実施するのが望ましいと考えられる。

小テストの形式は、穴埋め式、選択式、記述式などがある。どの方法も、学生がどの程度授業の内容を理解しているかを知ることができる。その中でも、記述

式は学生の理解の程度を知ることに対して最も適している。なぜなら、穴埋め式、選択式は授業の内容を深く理解していなくても正解することができる。それに対し、記述式は授業の内容を深く理解していなければ解答できず、間違った場合もどのように理解しているかもわかるからである。よって、本研究では、記述式小テストにより授業の内容の理解の程度を知る方法を取り扱う。

2.2 記述式小テストの解答の取り扱い

記述式の小テストで学生の理解の程度を把握することは穴埋め式、選択式に比べて講師の負担が大きくなる。1つの授業における学生の人数は数十人であり、この解答すべてに目を通し把握するには時間がかかるからである。これを解決するためには、解答をその内容に沿って整理することが有効である。

ここで、どのように整理するかが問題となる。小テストの解答の傾向を把握しやすくするためには全ての解答にどのような内容があるか、同じ内容の解答がいくつ存在するかという情報が必要になる。また、授業中にフィードバックを行うなら、各学生の理解の程度を把握してフィードバックするのは時間がかかり現実的ではないので、全学生の理解の程度を知るために同じ内容の解答が多い解答群の情報が必要となる。そして、個別にフィードバックを行うのであれば、各学生の傾向を知らなければならず同じ内容の解答が少ない解答群の情報も必要となる。本研究では授業中にフィードバックを行うことに焦点をあて、全学生の傾向を把握するために同じ内容の解答が多数ある解答群を把握できるようにする。

以上のことを実現するためには、解答を何らかの形で分類し、講師に提示することが有効である。分類を行う技術としてクラスタリングが知られている。これは、与えられたデータをある基準に従い、クラスタと呼ばれるいくつかのグループに分割する手法である。また、高次元（多くの内容を含む）データをデータ同士の距離を保存しながら次元を減らすことで低次元のデータに変換し表示する低次元変換法も多数ある解答の把握に有効である。

図 2.1 にクラスタリングとベクトルの低次元表現法のイメージを示す。

クラスタリングで解答を取り扱う方法では、各解答のベクトルの距離や類似度を測り、似た内容のものを同じクラスタに属するように分類する。この結果により講師はクラスタ内の解答を1つ見ることで、その内容を把握できる。そのため、クラスタの数の分だけ解答を見るだけで、すべての解答に目を通すことなく解答の傾向を把握できる。

ベクトルの低次元表現法では、元のベクトルの次元の距離をできるだけ保ったまま低次元で各解答の関係を表現する。これにより、同じ内容の解答が近くに、そうでない解答が遠くに配置されるので解答の内容がどの程度偏っているかを視覚的に把握することができる。また、講師の主観によるが、同じ内容の解答がある程度わかるので講師がすべての解答を読む必要がなくなる。

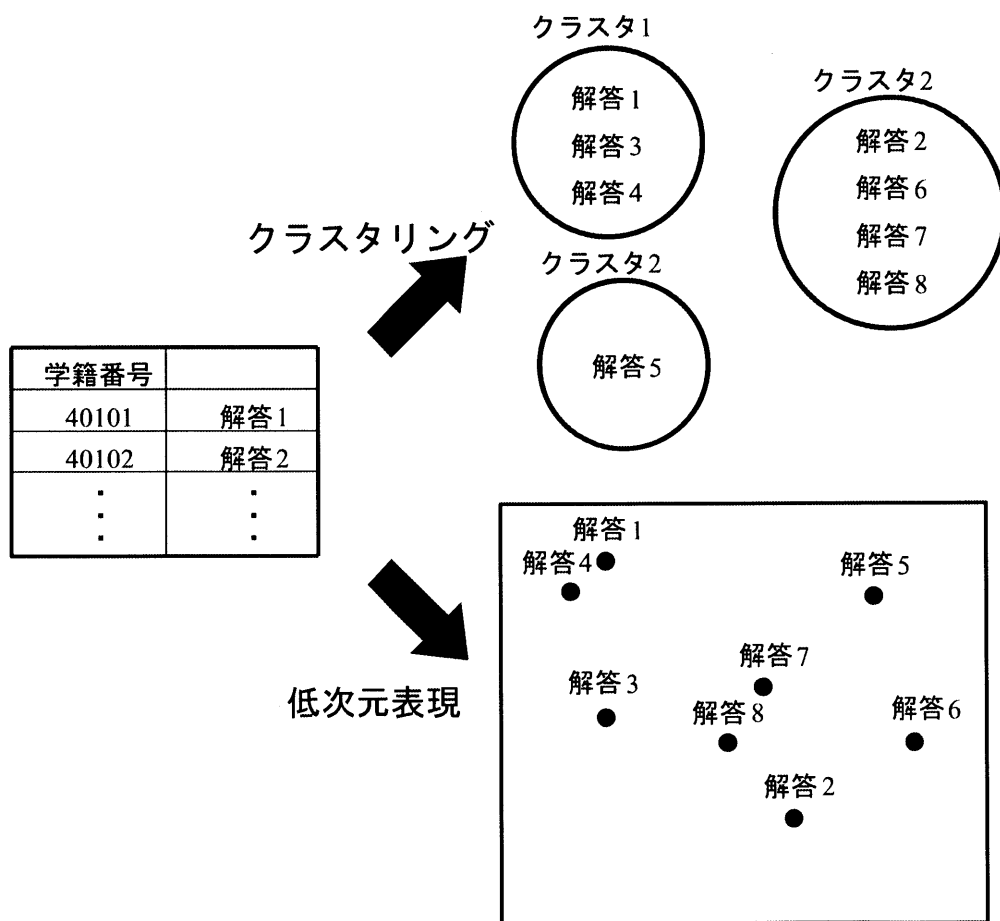


図 2.1: 解答の整理法のイメージ

第3章 文書処理

ここでは、文書のクラスタリングと低次元表現法の手法に共通した日本語で書かれた解答の処理の方法について述べ、クラスタリングや低次元表現法の代表的な手法について述べる。そして、クラスタリングの具体的な方法と低次元表現法の具体的な方法について述べる。

3.1 文書の特徴ベクトル

文書を扱いやすくするために、各文書から特徴ベクトルを作成する。特徴ベクトルは文書の内容を反映するように作成する。内容を表すために、文書中から適切な単語をキーワードとして抽出する。各キーワードに適切に重みを求めることで、文書の特徴ベクトルとして表現する。

ここで重要なことは、文書の特徴を引き出すためのキーワードの選び方と重みの決め方である。

そのため、どの単語をキーワードとするか、どのように重みをつけるかなど特徴ベクトルの作成法はさまざまな方法が提案されている [1][4][5]。本稿では、一般の文書分類で良く用いられている $tf \cdot idf$ 法と、記述式小テストの分類のために本研究で提案されている大井らの方法について説明する。

3.1.1 $tf \cdot idf$ 法

$tf \cdot idf$ 法は、特徴ベクトルの重みを決定する方法であり、キーワードの重要度を tf と idf の積で求める [5]。一般的に使用されるキーワードは名詞のみである。 tf (term frequency) は、ある文書で繰り返し使用されている単語ほど重要であるという考えであり、 idf (inverse document frequency) は全文書で繰り返し使用されている単語は重要でないという考え方に基づき定める重みである。その結果、ある文書で使用される回数が多い単語ほどその重みは大きくなり、全文書で使用し

ている文書が多い単語ほどその単語の重みは小さくなる。 tf は、ある文書 d 中に出現する単語 t の頻度であり、 $tf(t, d)$ で表す。 $idf(t)$ を単語 t が全文書でどれくらいの文書に出現するかを表す尺度で、例えば次式の $idf(t)$ で定義される。

$$idf(t) = \log \frac{N}{df(t)} + 1 \quad (3.1)$$

ここで、 N は全文書数、 $df(t)$ は単語 t が出現する文書数である。 $N/df(t)$ の対数をとるのは文書集合の規模と単語 t が使用されている数にに対して idf の変化を小さくするためである。

$tf \cdot idf$ は、 $tf(t, d)$ と $idf(t)$ を掛け合わせたものである。

$$tf \cdot idf = tf(t, d) \times idf(t) = tf(t, d) \left(\log \frac{N}{df(t)} + 1 \right) \quad (3.2)$$

例えば、対象となる全文書数が $N = 100$ 、キーワードとなる単語が全文書で使用されている文書数を表 3.1 に示す。このとき、次の文書をベクトル化する。

メールの内容に注意する。具体的には、
相手が傷つかないように文章に注意してメールを送る。

表 3.2 に tf 値、式 (3.1) で計算した idf 値、そして $tf \cdot idf$ 値を示す。 tf 値について見ると、「メール」という単語が 2 回使用されているのでその値は大きく、「相手」という単語が使用されている単語が 1 回なのでその値は小さくなっている。 idf 値は、キーワードの使用回数が多いとその値は小さく、使用回数が少ないとその値は大きくなっている。そして、これらの値を掛け合わせた $tf \cdot idf$ 値をこの文書の特徴ベクトルとする。 $tf \cdot idf$ 値は、ある文書でキーワードとなる単語が使用されている回数が多ければその値は大きく、使用されている回数が少なければ小さくなっており、全文書でキーワードとなる単語が使用されている文書数が多ければその値は小さく、使用されている文書数が少なければその値が大きくなっている。

表 3.1: キーワードとなる単語が使用されている文書数

キーワード	メール	人	相手	内容	注意	文章	具体的
使用回数	80	50	30	10	7	5	3

表 3.2: tf・idf法の各値

キーワード	メール	人	相手	内容	注意	文章	具体的
tf 値	2	0	1	1	2	1	1
idf 値	1.10	1.30	1.52	2.00	2.15	2.30	2.52
tf・idf 値	2.20	0	1.52	2.00	4.30	2.30	2.52

3.1.2 大井らの方法

tf・idf法を代表とする一般的な特徴ベクトルの作成法は、大規模な文書群を対象としている。大規模な文書群とは、何万件もの文書で、各文書は数十行から数百行であるようなものを指す。それに対して、本研究で対象としている文書は授業での小テストの解答であり、高々100件程度で各文書は高々数行である。そのため、一般的な特徴ベクトルの作成法をそのまま用いたのでは、分類の性能が悪くなってしまう [1]。

この問題を解決するために、大井らはこのような小テストの解答から、的確にその文書の内容を抽出できる特徴ベクトルの作成法について検討した [1]。この手法により生成した特徴ベクトルを用いることで、精度の高い分類を行うことができる。本研究でも、これを用い解答の特徴ベクトルを作成する。

この手法は、特徴ベクトルの対象の品詞の拡大と、係り受けに基づく重み付けに基づいている。ベクトル化の対象の品詞の拡大では、文書の主要構成要素が主語（名詞）と述語（動詞）であるという考えから、キーワードの対象となる品詞を名詞と動詞（語幹のみを用いる）としている。また、係り受けに基づく重み付けは、より多くの文節から（直接的、間接的に）修飾を受けている文節ほど、その文書の内容をうまく特徴付けていると考え、多く修飾されている文節に含まれているキーワードを重視する。

この方法の手順を具体的に示す。

1. 文書中の各文を係り受け解析器 [6] で解析する。
2. 各文節について、被修飾回数を1とする。
3. 修飾されていない文節から順に、その文節の被修飾回数を、その文節が修飾している文節の被修飾回数に加算する。
4. 一文ごとに、各文節の被修飾回数を [0.5, 1] に正規化する。

表 3.3: 大井らの手法での各キーワードの重み

キーワード	知る	人	添付ファイル	注意
重み	1	2	3	5
正規化後の重み	0.50	0.63	0.75	1.00

5. 各文書から、動詞と名詞をキーワードとして抽出し、その重みを、それが属している文節の正規化した被修飾回数とする。

例えば、次の文書が大井らの方法で特徴ベクトル化する方法を説明する。

知らない人からの添付ファイルには特に注意する。

始めに、係り受け解析器でこの文書で解析すると図 3.1 のようになる。次にすべての文節の被修飾回数を 1 する。「人から」という文節は、「知らない」という文節から修飾されているので被修飾回数は 2 となる。「添付ファイルには」という文節の被修飾回数は、「人から」のという文節に修飾されているので「人からの」の被修飾回数を足して 3 となる。このように、ある文節の被修飾回数は、ある文節を修飾している文節の被修飾回数を加えることで決定する。この文書の重みをすべて決定すると図 3.2 のようになる。最後に、名詞と動詞をキーワードとして抽出し、その重みを被修飾回数とし、 $[0.5, 1]$ に正規化する (表 3.3)。

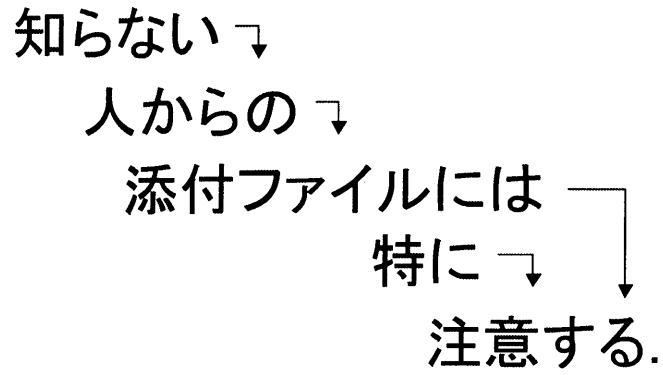


図 3.1: 係り受け解析器での解析結果

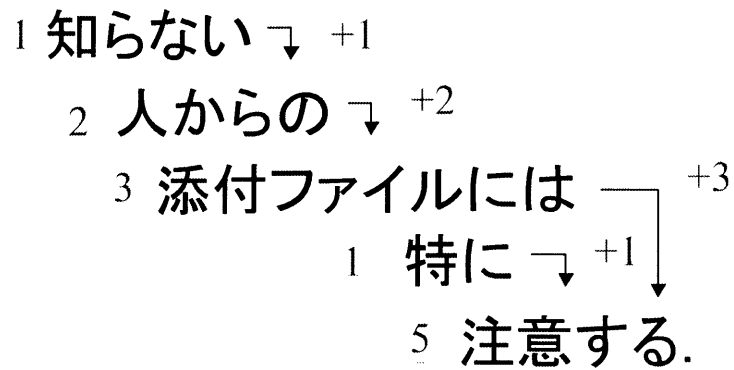


図 3.2: 大井らの方法の各文節の重み

3.2 文書分類における多変量解析

多変量解析とは、多数の変数からなる情報（今回は特徴ベクトル）を分析し、有用な情報を抽出する手法である。主成分分析、重回帰分析などがよく知られているが、文書処理には適さない。これらの手法は、すべての特徴ベクトルの次元の数を同じ基準で削減し分析を行う。しかし、文書分類では文書群がどのような内容を含んでいるのかを知ることが目的としており、内容ごとに重要な次元が異なるため同一の基準で次元数を削減することは無理がある。そこで文書分類では、特徴ベクトル間の距離に基づいて次元数を削減し、分析する手法がよく用いられる。

本節では、多変量解析の代表的な手法としてワード法を用いた階層型クラスタリング、 k -means 法による非階層型クラスタリング、多次元尺度法、そして自己組織化写像について説明する。なお、以下これらの手法を特徴ベクトル間の距離に基づく分析手法と呼ぶ。また、特徴ベクトル間の距離とは特徴ベクトル同士の類似度を意味するものとする。具体的には、ユークリッド距離、余弦距離などが用いられることが多い。

3.2.1 階層型クラスタリング

階層型クラスタリングは、各入力データの距離を計算し近いものを1つにまとめ、さらにまとめられたもの同士の距離を測り近いものを1つにまとめることを繰り返すことでクラスタリングを行う方法である。多くの場合、その結果はデンドログラムとして表示する。

デンドログラムとは、各終端ノードが各入力データを表し、併合されてできたクラスタを非終端ノードで表した木である。デンドログラムの例を図3.3に示す。例の非終端ノードの高さは、併合されたときのクラスタ間の距離を表す。下にある分岐点ほどより類似したデータを併合したことを表す。適当な高さで分断することで、任意のクラスタ数に分類できる。例えば、図3.3において点線の位置で分断することで、4個のクラスタを得る。

階層型クラスタリングは、距離の求め方の違いによりさまざまな方法がある。ここでは、よく使用されているワード法について説明する。

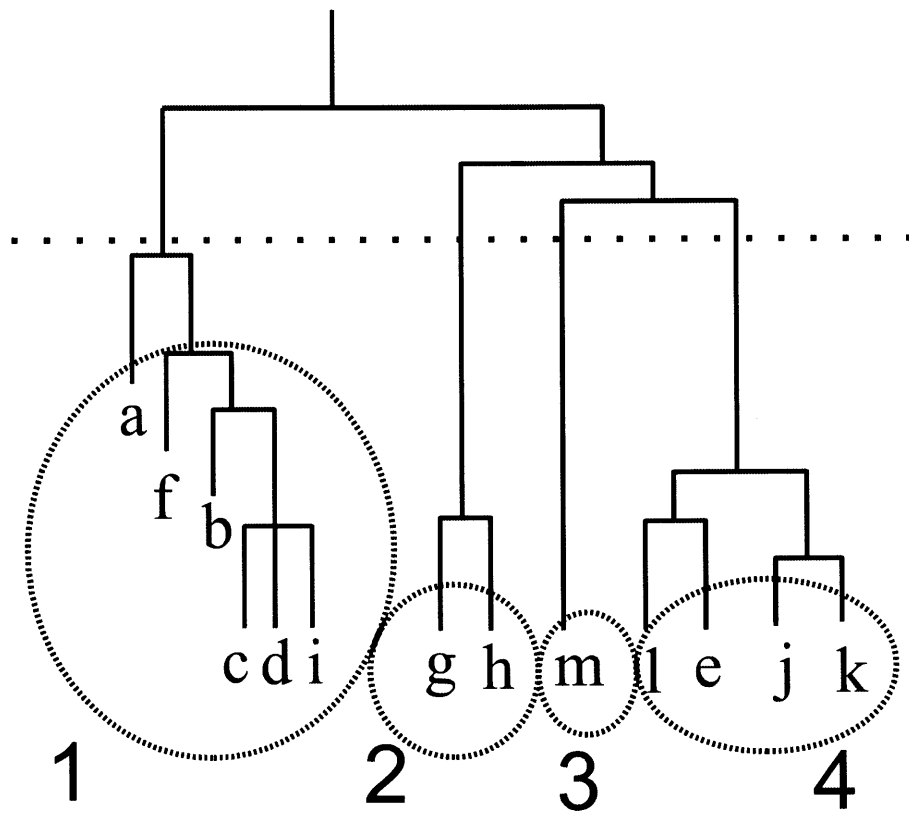


図 3.3: デンドログラムの出力例

ウォード法

ウォード法は、各入力データとクラスタの重心との距離の2乗値の和をできるだけ小さくするように、クラスタの併合による2乗和の増分が最も小さいクラスタ同士を併合する [7]. 与えられるデータを $\mathbf{x}_1, \dots, \mathbf{x}_n$ ($\mathbf{x}_i = (x_1, \dots, x_d)$), 分割されたクラスタを C_1, \dots, C_k , あるクラスタ C_j について, C_j に含まれる入力データと C_j の重心との距離の2乗値の和 $E(C_j)$ を

$$E(C_j) = \sum_{\mathbf{x}_i \in C_j} \|\mathbf{x}_i - M(C_j)\|^2 \quad (3.3)$$

とする. ここで, $M(C_j)$ は, クラスタ C_j の重心で

$$M(C_j) = \frac{1}{|C_j|} \sum_{\mathbf{x}_i \in C_j} \mathbf{x}_i \quad (3.4)$$

である. 次にウォード法でクラスタリングする手順を示す.

1. 1個の入力データだけを含む n 個のクラスタがある初期状態とする.
2. クラスタ C_l, C_m を併合したときの $E(C_l, C_m)$ の増分を計算する.

$$\Delta E(C_l, C_m) = E(C_l \cup C_m) - E(C_l) - E(C_m)$$

3. $\Delta E(C_l, C_m)$ が最小となる二つのクラスタを逐次的に併合する. 全ての入力データが一つのクラスタに併合されたら終了. そうでなければステップ2へ.

3.2.2 非階層型クラスタリング

非階層型クラスタリングは、分割の良さの評価関数を定め、その評価関数を最適にする分割を探索する. 本稿では、代表的な k -means 法について説明する.

k -means 法

k -means 法は、対象となるデータ群を与えられたクラスタ数 k 個に、分割する手法である [8]. この手法では、各クラスタを代表する点 (代表点) はクラスタの平均であり、次式の評価関数を最小にする.

$$\sum_{i=1}^k \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \mathbf{c}_i\|^2 \quad (3.5)$$

ここで、 \mathbf{x} は入力ベクトル、 c_i はクラスタ C_i に属する入力の平均である。クラスタリングの手順は以下のとおりである。

1. k 個の代表点を乱数で決定する。
2. 各入力データをそれに最も近い代表点に割り当てる。
3. 各代表点について、それに割り当てられた入力データの各要素の重心を求め、その代表点を更新する。
4. もしステップ2で、すべての代表点への割り当てが、以前の割り当てから変化しなかったなら終了。そうでなければステップ2へ戻る。

3.2.3 低次元表現

低次元表現は、多次元のベクトルのデータを低次元の空間（2次元の平面）上に視覚的に表示することで、データ構造を洞察しやすくするための方法である。ここでは、ベクトル間の距離に基づいて低次元に表示する方法である多次元尺度法と自己組織化写像について説明する。

多次元尺度法

多次元尺度法は、ある高次元データの構造を低次元空間で表示させ、その低次元空間での点間の距離を元の空間での相違性に関連付ける方法である [9]。

$\mathbf{y}_i (i = 1, \dots, n)$ を $\mathbf{x}_i (i = 1, \dots, n)$ の低次元像、 δ_{ij} を \mathbf{x}_i と \mathbf{x}_j の距離、 d_{ij} を \mathbf{y}_i と \mathbf{y}_j の距離とする。このとき、 $n(n-1)/2$ 個の像点間の距離 d_{ij} を元の空間の距離 δ_{ij} に近づくように \mathbf{y}_i を定める。

すべての i, j に対して $d_{ij} = \delta_{ij}$ となる \mathbf{y}_i を見つけることは通常できないので、ある評価関数を設けてそれを最小にすることで \mathbf{y}_i を求める。以下に誤差2乗和関数の評価関数の例をいくつか示す。

$$J_{ee} = \frac{\sum_{i < j} (d_{ij} - \delta_{ij})^2}{\sum_{i < j} \delta_{ij}^2} \quad (3.6)$$

$$J_{ff} = \sum_{i < j} \left(\frac{d_{ij} - \delta_{ij}}{\delta_{ij}} \right)^2 \quad (3.7)$$

$$J_{ef} = \frac{1}{\sum_{i < j} \delta_{ij}} \sum_{i < j} \frac{(d_{ij} - \delta_{ij})^2}{\delta_{ij}} \quad (3.8)$$

これらの内いずれかの評価関数を用い、それを最小にすることである高次元データの構造を低次元空間で表示させることができる。それぞれの特徴は、 J_{ee} は、大きな積算誤差を強調し、 J_{ff} は大きな誤差率を強調する。そして J_{ef} は大きな誤差と誤差率の損を強調する。

用いる評価関数を決めた後、その評価関数を最小にする \mathbf{y}_i を決定すればよい。これは、最急降下法によって評価関数を最小にする \mathbf{y}_i を求めることができる。最急降下法は、ある初期配置を定め、 \mathbf{y}_i を評価関数の減少が最も大きくなる方向に変化させる操作を繰り返すことで最小値を求めるものである。したがって、式 (3.6) から式 (3.8) の \mathbf{y}_i に関する勾配を求め \mathbf{y}_i を更新すればよい。低次元空間において距離は $d_{ij} = \|\mathbf{y}_i - \mathbf{y}_j\|$ で、 \mathbf{y}_i に関する d_{ij} の勾配は $\mathbf{y}_i - \mathbf{y}_j$ 方向の単位ベクトルになるので式 (3.6) から式 (3.8) の勾配はそれぞれ次のようになる。

$$\nabla_{\mathbf{y}_k} J_{ee} = \frac{2}{\sum_{i < j} \delta_{ij}^2} \sum_{i \neq j} (d_{kj} - \delta_{kj}) \frac{\mathbf{y}_k - \mathbf{y}_j}{\delta_{kj}} \quad (3.9)$$

$$\nabla_{\mathbf{y}_k} J_{ff} = 2 \sum_{j \neq k} \frac{d_{kj} - \delta_{kj}}{\delta_{kj}^2} \frac{\mathbf{y}_k - \mathbf{y}_j}{d_{kj}} \quad (3.10)$$

$$\nabla_{\mathbf{y}_k} J_{ef} = \frac{2}{\sum_{i < j} \delta_{ij}} \sum_{k \neq j} \frac{(d_{kj} - \delta_{kj})}{\delta_{kj}} \frac{\mathbf{y}_k - \mathbf{y}_j}{d_{kj}} \quad (3.11)$$

自己組織化写像 (Self Organizing Map)

Self Organizing Map (以下 SOM) は、入力データの類似度を反映したマップを生成するニューラルネットワークの教師なし学習法である [10]。構造は、入力層とマップ層の2層からなり (図 3.4)、マップ層は2次元六方格子上に等間隔に配置されたノード群で構成される。

SOMの動作を簡単に記す。SOMの入力は V 次元のベクトル $\mathbf{x} = [x_1, x_2, \dots, x_V]$ である。SOMの各ノードは、入力と同じ次元を持つ参照ベクトル m_i をもつ (i はノード番号とする)。SOMの出力は、入力に最も近い参照ベクトルを持つノードである。すなわち、 \mathbf{x} に最も近い参照ベクトルが m_j であったとすると、第 j ノードの出力を 1 とし、他を 0 とする。以下この状態を、「入力 x をノード j に分類した」と表現する。参照ベクトルを学習により適切に定めることで、距離が近い入力同士を近くのノードに、遠いものを遠くのノードに分類することができる。

SOMの学習は、類似する入力ベクトルを、六方格子上で近いノードに分類するように、各ノードの参照ベクトルを調整する。学習するデータとして、入力ベクトルの集合 $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ が与えられたとする。時刻 t に入力データ $x(t) = \mathbf{x}_i$ ($1 \leq i \leq N$)

が入力されたとき、次式に基づきノードの参照ベクトルを更新する。

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{c(\mathbf{x}),i(t)}[\mathbf{x}(t) - \mathbf{m}_i(t)] \quad (3.12)$$

$$c(\mathbf{x}) = \arg \min_i [\|\mathbf{x}(t) - \mathbf{m}_i(t)\|] \quad (3.13)$$

ここで、係数 $h_{c,i}$ は、近傍関数と呼ばれ、ガウス関数が良く用いられる。

$$h_{c,i} = \alpha(t) \exp\left(-\frac{(\|\mathbf{r}_i - \mathbf{r}_c\|)^2}{2\sigma^2(t)}\right) \quad (3.14)$$

ここで、 \mathbf{r}_i は2次元格子状に配置された第 i ノードの座標を表し、 $\alpha(t)$ と $\sigma^2(t)$ は、学習を制御するパラメータである。学習は、設定回数に達するか、参照ベクトルが収束すれば終了となる。本研究では、式 (3.12) と式 (3.13) で示したように、各サンプルが提示される毎にパラメータを更新するオンライン学習法ではなく、すべてのサンプルに対する応答を求めた後、参照ベクトルを一括して更新するバッチ学習を用いる。参照ベクトルの更新として以下の式を用いる [11]。

$$m_i = \frac{\sum_j n_j h_{j,i} \bar{x}_j}{\sum_j n_j h_{j,i}}, \quad \bar{x}_j = \sum_l \frac{x_l}{n_i} \quad (3.15)$$

ただし、 l は $c(\mathbf{x}_l) = i$ となるインデックスであり、 n_i は $c(\mathbf{x}_l) = i$ となるサンプル数を表す。

以上のように与えられたデータから参照ベクトルを更新し学習を行う。学習後、各入力ベクトルを SOM に入力し一番近い参照ベクトルを持つノードに割り当て、それらをまとめて表示することで図 3.5 のようなマップを得る。

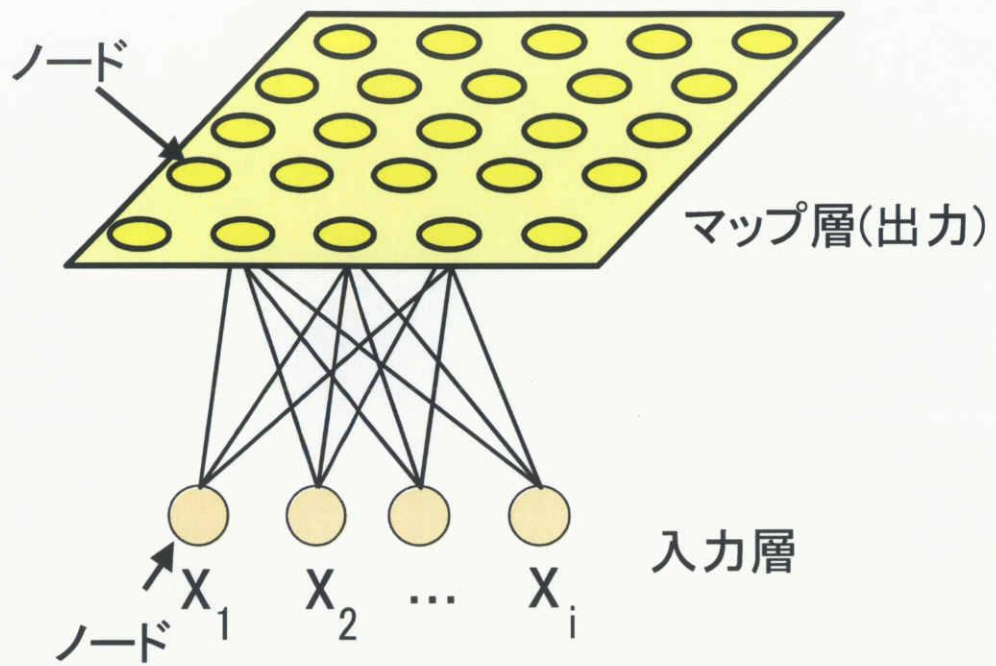


図 3.4: SOM の構造

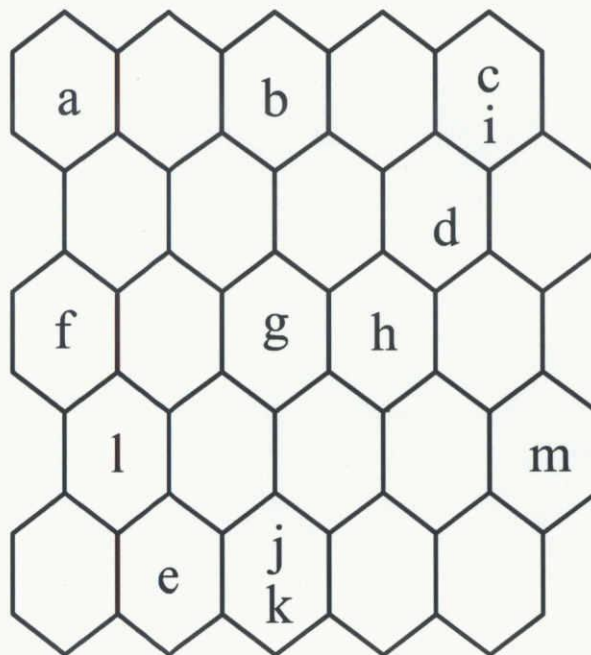


図 3.5: SOM の出力例

第4章 提案法

この章では、前章で説明した各手法について比較検討を行い、講師が小テストの解答の傾向の把握を素早く行うのを補助するために、必要となる手法について述べる。

そして、提案法である分析方法にSOMを採用し、さらにSOMの結果に k -means法を用いる方法と、クラスタの内容がわかるように各クラスタの代表の解答を選出する方法について述べる。

4.1 記述式小テストの解答の傾向を把握するための分類

小テストの解答を自動分類するにあたり、要求される事柄は、

1. 類似した内容の解答が1つにまとまっている。
2. クラスタ間の関係が視覚的にわかる。
3. クラスタに属する解答の主要な内容がわかる。

である。例を図4.1に示す。この図は、類似した内容がクラスタという形で1つにまとまっており、クラスタ間の内容の違いを反映させている。つまり、クラスタ1とクラスタ2は近い位置に配置されているので、この2つのクラスタの内容が類似しており、クラスタ1とクラスタ3は遠い位置に配置されているので、この2つのクラスタは違った内容であるというように視覚的にクラスタ間の関係がわかる。そして、各クラスタの内容が分かる情報を表示している。

これは分類の目的が、採点ではなく、講義へのフィードバックを速やかに行うための補助であることに起因している。講義へのフィードバックを行うためには、個々の解答がどのように分類されるのかではなく、解答群にどのような内容がどのような割合で含まれているのかを知ることが重要である。

4.1.1 各手法による解答の分類

3.2節で説明したように、一般に分析手法の代表的なものに、ワード法などの階層型クラスタリング手法（結果はデンドログラムで表示）、 k -means法があり、データ構造を見やすくするために多次元データを2次元または3次元に表示する方法として多次元尺度法とSOMがある。ここでは、各手法について比較検討を行い、どの手法が4.1節の要求を実現できるかについて述べる。

デンドログラム

デンドログラムで小テストの解答の分類結果を表示することを考える。デンドログラムの出力の例を図4.2に示す。図の a, b, \dots, m は、データを表す。低い高さでまとまっている枝ほどより類似している。解答の類似度を、各枝のまとまり方で見ることができると主要なクラスタを知ることはできる。例えば、4.2では、一番低い高さでまとまっているデータ j と k が一番類似度が高い。もう少し大きいまとまりで見ると、データ l, e, j, k が類似度が高いとわかる。

しかし、それらがどのような内容を持っているのかは知ることができない。なぜなら、デンドログラムは、各解答（より小さなクラスタ）間の距離に基づいて作成したものであり、各クラスタの内容が何であるのかという情報が残されていないためである。そのため、主要な内容を把握するという今回の要求を満たさない。

k -means法

2次元のデータ群に対する k -means法の出力の例を図4.3に示す。図の a, b, \dots, m の点は2次元平面状のデータの位置を、*は各クラスタの代表点を、直線は境界線を表す。各クラスタの代表点が求められているため、デンドログラムの場合と異なり、これを用いることで各クラスタに分類された解答の平均的な内容を知ることができる。しかし、分類対象のデータが多次元になると、各データの相互関係を適切に図示することができない。そのため、与えられたデータがどのクラスタに分類されたのかという情報と、各クラスタの代表点の情報しか提供されない。クラスタの相互関係は、平均的な内容の距離を測ればわかるが、距離の表示は表で表示しなければならずクラスタの数が多くなると見にくくなり講師に負担がかかってしまう。そのため、今回の使用には適さない。

多次元尺度法

多次元尺度法の出力の例を図4.4示す。図の a, b, \dots, m は、データを表す。この方法では、2次元平面上に類似した内容の解答は近く、そうでない解答は遠くなるように表示される。そのため、二つの内容を持つあいまいな解答も対応する二つの解答の中間付近に表示されるため、解答同士の関係もわかりやすい。しかし、解答同士の関係が把握できるが、クラスタが明示的に形成されない。また、何らかの手法でクラスタを得られたとしても、この方法の適用の結果得られた情報が低次元上の点だけであるため、クラスタの内容を知ることができない。

SOM

SOMの出力の例を図4.5に示す。図の a, b, \dots, m は、各解答を表す。SOMは、 k -means法と同様に、各ノードの参照ベクトルからそのノードに分類された解答の内容を知ることができる。また、多次元尺度法と同様に、分類結果は、一般に2次元のマップ上に、類似した内容の解答は近く、そうでない解答は遠くなるように表示される。そのため、解答同士の関係が把握できる。また、各ノードを一つのクラスタと見ることでクラスタも形成できたと言える。よって、講師が記述式小テストの解答の傾向を把握しやすい方法であると言える。

4.2 提案する分類法

本研究では、これらの手法の特徴をふまえ、まずSOMで分類を行い、その結果得られた各ノード（参照ベクトル）を k -means 法で分類する手法を提案する。

SOMの問題点は、分類結果が細かいクラスタになってしまう点である。2次元のマップ上でより大きなクラスタを見いだす方法としてU-matrix法による表示がある[12]。U-matrixは、SOMの隣接するノードの参照ベクトル間の類似度を表す指標である。U-matrix法は、この指標を用いて、隣接するノード間を色分けすることで視覚的に表示する方法である。しかし、U-matrix法では、隣接するノードの関係が分かるが、離れたノードの関係は分からない。そのため、明確にクラスタが形成できない。本研究では、SOMの分類結果が細かいクラスタになってしまう問題点を k -means 法によって補う。 k -means 法をSOMの参照ベクトルに適用することで、SOMの分類結果を自動的に併合し、人が一目見てわかるぐらいの数である数個のクラスタにまとめることができる。具体的には、SOMの参照ベクトルを k -means 法で分類する。その結果、SOMで生成したマップ上に大まかに分類したときのクラスタ情報を付加することができる。

以下に、提案する分類の手順を記す。

1. 各解答の特徴ベクトルを作成する（大井らの方法）。
2. SOMにより、すべての特徴ベクトルを学習する。学習終了後、学習したSOMですべての解答を分類し、SOMマップとして表示する。
3. SOMのすべての参照ベクトルに対して、 k -means 法を適用する。分類終了後、SOMマップの各ノードを k -means 法の分類結果に従い区分けする。

講師は、 k -means 法の結果から、大まかな傾向を知ることができ、SOMの結果からクラスタの相互関係を知ることができる。その結果、すべての解答を見ることなく、解答の傾向を知ることができる。図4.6にSOMの参照ベクトルに k -means 法を適用した時のイメージ図を示す。この図の見方は、六方格子がSOMのノードを表し、その中にあるアルファベットが各入力を表している。図の色分けは、SOMの参照ベクトルに k -means 法を適用した結果、参照ベクトルがどのように分類されたかを示している。1つの色が1つのクラスタに対応している。これより、 j と k は極めて類似した内容を持ち、 f , l , e , j , k が比較的類似した内容を持つことがわかる。

全解答に含まれる内容の全種類数が k -means 法での k の値と比べ少ない場合、同じ内容の解答が複数のクラスタに分割されてしまうが、 k が十分に小さければ（10 未満）講師は容易にそのことに気づくことができる。

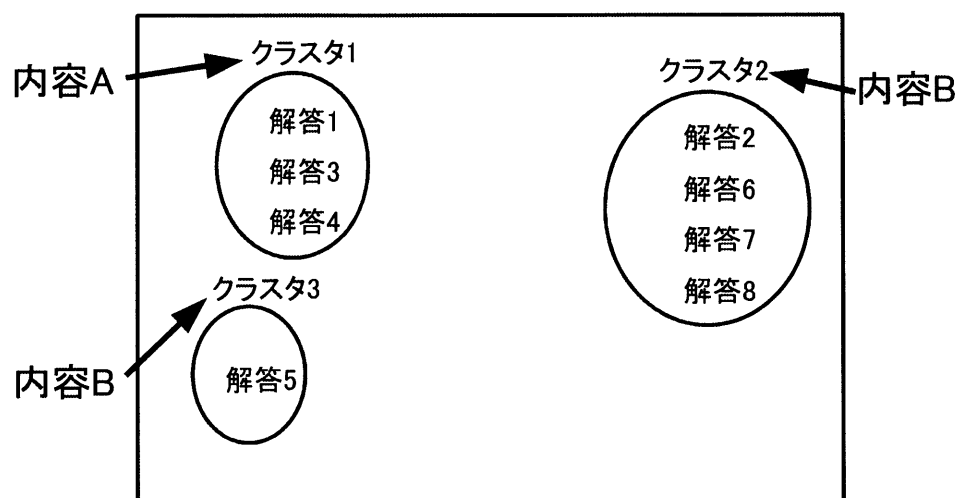


図 4.1: 理想的な分類結果

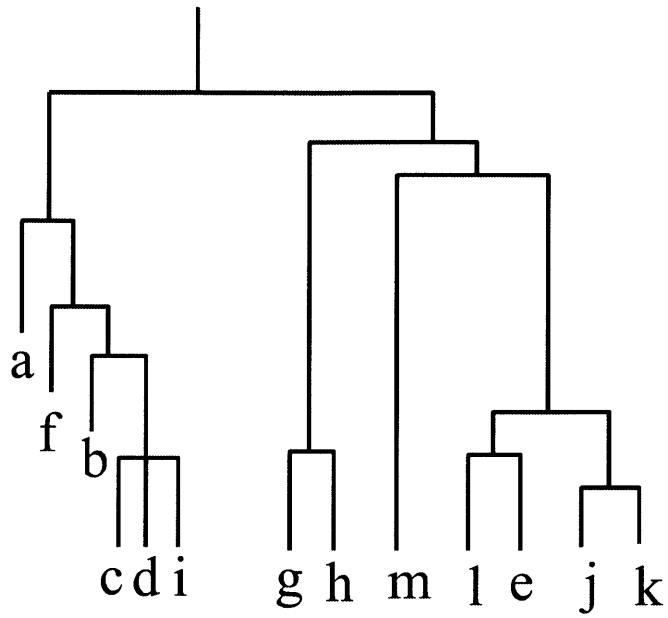


図 4.2: デンドログラムの出力例

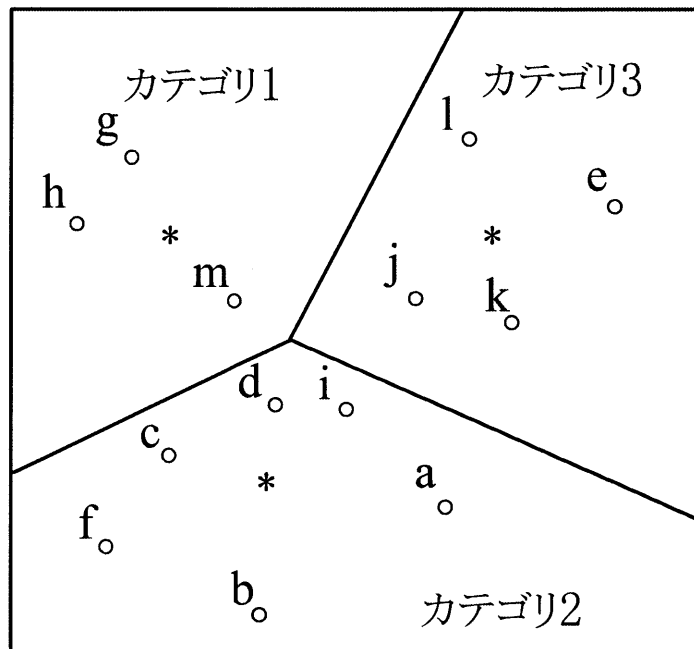


図 4.3: k -means 法の出力例

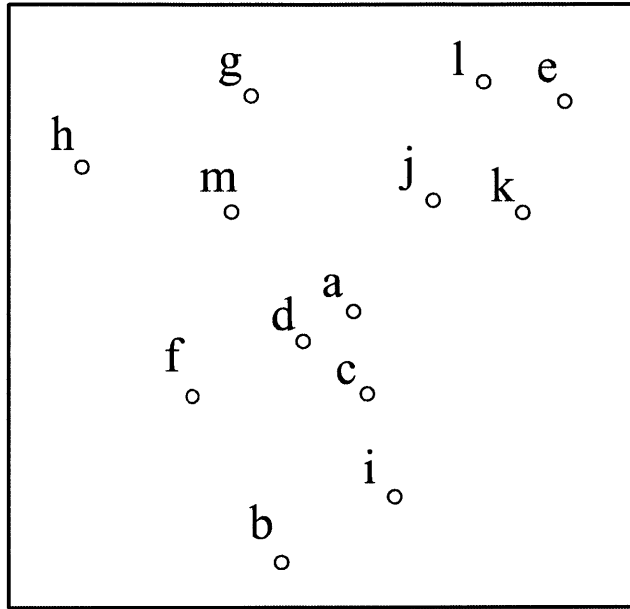


図 4.4: 多次元尺度法の実出力例

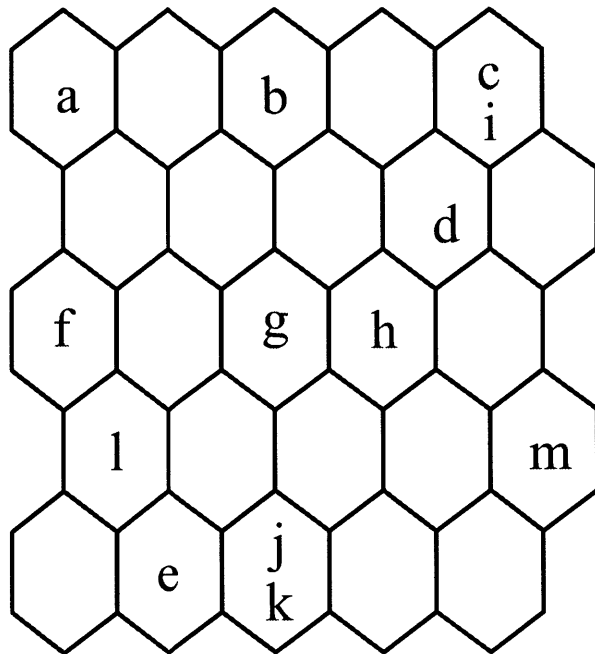


図 4.5: SOM の出力例

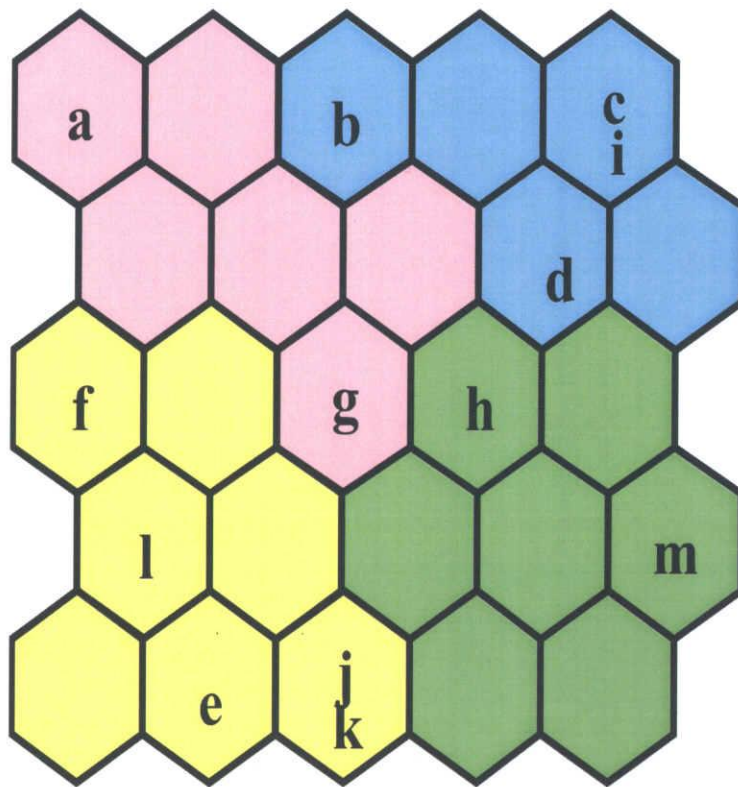


図 4.6: SOM の参照ベクトルに k -means 法を適用した結果

4.3 クラスタの内容の表示

4.2節で提案した手法を用いることで、解答を大まかに分類することができる。しかし、分類の結果得られた各クラスタの内容は分からない。そこで、その内容を分かりやすく表示する方法について、本節で検討する。

多くの場合、SOMや、 k -means法の分類結果の特性を生かして、クラスタの内容を表す単語をその代表点や参照ベクトルから抽出することが行われている [2]。しかし、小テストの解答は、ある問題の解答という意味ではよく似たものとなる。そのため、数個の単語だけでは、その内容を十分に把握できない。例えば、あるクラスタの内容を示すような単語が以下のように選ばれたとする。

インターネット, メール, 相手

これにより、インターネット、メールに関することを述べていることが分かる。しかし、相手という単語からはさまざまな内容が予想されるのでクラスタの内容を決定できない。そこで、このように抽出した単語ではなく、各クラスタ内からそれを代表する解を選択することで、各クラスタの内容を表示する。

同じクラスタ内に分類された解答でも、その内容にはある程度のばらつきがある。そこで、 k -means法の分類結果だけでなくSOMの分類結果にも着目する。クラスタの代表として、最も多くの解答が割り当てられているノードに着目する。SOMにより同じノードに割り当てられた解答は極めて類似した内容である。つまり1つのノードに割り当てられている解答が多いということは、非常に似た解答が多数あることを意味する。これは、平均的な内容を選ぶよりも同じ内容の解答が多いものを選んだ方が、多いものに注意を引かれるという人の感性に近いと考えたからである。選ばれたノードには複数の解答が分類されているが、見やすさを考慮して、この中から短い解答を選出するのが望ましい。しかし、短かすぎる解答の特徴ベクトルの情報量は少ないためSOMが誤分類しやすい。そこで、そのノードに割り当てられている解答の中で中間の長さの解答を選ぶ。以下にクラスタの代表の解答の選出方法の手順を示す。

1. 対象とするクラスタで、最も多くの解答が分類されたノードを探す。
2. もし、そのようなノードが複数ある場合は、それらの内で k -means法で得られた代表点に最も近い参照ベクトルを持つノードを選択する。

3. 探し出したノードに分類された解答のうち，中間の長さの解答を選択し，そのクラスタの代表とする．

なお，手順2は最も多くの解答が分類されたノードが，多数存在する場合の例外処理である．その場合は，平均的な内容を表す k -means 法の代表点に最も近い参照ベクトルを持つノードを選択する．

第5章 実験

本章では、提案法の有効性を実際に講義で行った記述式小テストの解答を用いて検証する。5.1節では、実験条件を述べ、5.2節では、本研究の提案法の結果を示し、考察を行う。

5.1 実験条件

実験については、提案法に実際の授業で行った記述式小テストの解答を用いて行う。対象としている講義は、三重大学工学部電気電子工学科1年生を対象に開講された2006年度の講義「計算機基礎I及び演習」である。表5.1に用いた解答のデータの情報を示す。また、付録Aに全解答の全文書を示す。小テストの解答の特徴ベクトルの作成に用いた方法は大井らの方法である。そして、特徴ベクトル間の距離にはユークリッド距離を用いた。また、表中のアルファベットは人手で小テストの解答を分類したときのクラスタのラベルである。ラベルEの「その他」とは、複数の内容を含むものと同じ内容が1人しかいない解答である。

学習の条件は、SOMの参照ベクトルの初期値をランダムとし、SOMのマップの大きさは入力データ数である解答数より少し多くなるように 10×9 とした。また、式(3.14)の σ は、初期値を $\sigma_{ini} = 3.0$ 、最低値を $\sigma_{fin} = 1.0$ とし、次式により学習回数に比例減少するように設定した。

$$\sigma(t) = \sigma_{ini} - \frac{t-1}{T-1} \times (\sigma_{ini} - \sigma_{fin}) \quad (5.1)$$

ここで、 T は最大学習回数であり、 t は現在の学習回数である。なお、学習はバッチ学習により行った。本実験では、最大学習回数を11とした。

k -means法では、あらかじめクラスタ数を設定しなければならない。実際のクラスタ数は、事前にわからない。そこで、クラスタ数を8と定めた。これは人が一目でわかる範囲で、できるだけ多いクラスタ数として選んだ。

なお、実験はCeleron (2.53GHz) 504MRAMにおいてMatlab, version7.2.0を用いて行った。また、SOMの結果の表示はSOM Toolbox[12]を用いた。

表 5.1: 実験に用いたテストの問題とその解答の概略

問題文	インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か?
出題時期・有効回答数	2006年度前期, 有効回答数 86
ベクトルの次元	186
主要な内容とその人数	A プライバシーの保護 (10人)
	B 法律 (著作権・肖像権) (35人)
	C 言葉づかい (8人)
	D 公開されている情報の正確さ (28人)
	E その他 (5人)

5.2 実験結果

図 5.1 に、SOM による小テストの解答の分類結果 (SOM の参照ベクトルに k -means 法を適用した結果) の一例を示す。図 5.1 の、色分けは k -means 法により分割されたクラスタを、大きく描かれた数字はクラスタの番号を表している。また、六方格子はノードを表し、その中に書かれている各 3 桁の数字は、そのノードに分類された解答の番号を表す。また表 5.2 に、提案手法により各クラスタから選ばれた代表の解答を示す。図 5.1, 表 5.2 が本研究の提案法の出力である。この 2 つの図表を見ることで、講師は記述式小テストの解答の傾向を把握する。また参考のため表 5.3 に各クラスタの代表の解答として選ばれた解答の内容を表すラベルを示す。

本研究では、同じ内容の解答が多数ある解答群について傾向を知りたい。そのため、人手で分けたラベルで人数が多いラベル B とラベル D について、情報検索の正確さを表す指標である再現率と適合率 [3] を計算した。

再現率とは、同じ内容の解答 (人手で分けたときのラベルが同じ解答) がどの程度抽出されるかということを表す指標で、以下のように定義される。

$$\text{再現率} = \frac{\text{結果から選ばれた解答の中で同じ内容 (ラベル) の解答の数}}{\text{全解答の中での同じ内容 (ラベル) の解答の数}} \quad (5.2)$$

一方、適合率とは結果から選ばれた解答の中で同じ内容 (ラベル) の解答がどの程度の割合で存在するかということを表す指標で、以下のように定義される。

$$\text{適合率} = \frac{\text{結果から選ばれた解答の中で同じ内容 (ラベル) の解答の数}}{\text{結果から選ばれた解答数}} \quad (5.3)$$

ここで、式 (5.2)、式 (5.3) の「結果から選ばれた解答」とは、表 5.2 に記された代表の解答の内容が、指定されたラベルと同じであるクラスタに含まれている解答のことである。また、「結果から選ばれた解答の中で同じ内容 (ラベル) の解答」とは、それらの解答のうち指定されたラベルと同じ内容を持つものを表す。これらの指標を計算することで、提案法がどの程度の分類精度を持っているかがわかる。再現率が高ければあるラベルの内容の解答をほとんど抽出できていることになり、適合率が高ければあるクラスタに他の内容のものがほとんど混ざっていないことになる。理想的には両者の値を 1 に近づけることが望ましいが、実際にはこれらの関係はトレードオフの関係が存在し、再現率を上げようとすれば適合率が下がり、適合率を上げようとすれば再現率が下がる。

提案法で用いる SOM と k -means 法は、初期値に乱数を用いるので、10 回の実験を行い、その平均を記す。表 5.4、表 5.5 に各ラベルについて計算した再現率と適合率を示す。

また、各クラスタ内の最頻数であるラベルが代表の解答のラベルと一致しているかを見るための指標として、指定された内容の解答が最頻出となったクラスタのうち、クラスタ内で最頻出となった内容とそのクラスタの代表の解答の内容が一致している割合を、一致率 R として求めた。この値が高ければ、各クラスタで多数を占める内容の解答を正しくピックアップできているということになる。

表 5.6 に、10 回の違った SOM と k -means 法の初期値で学習を行った結果をそれぞれ示す。

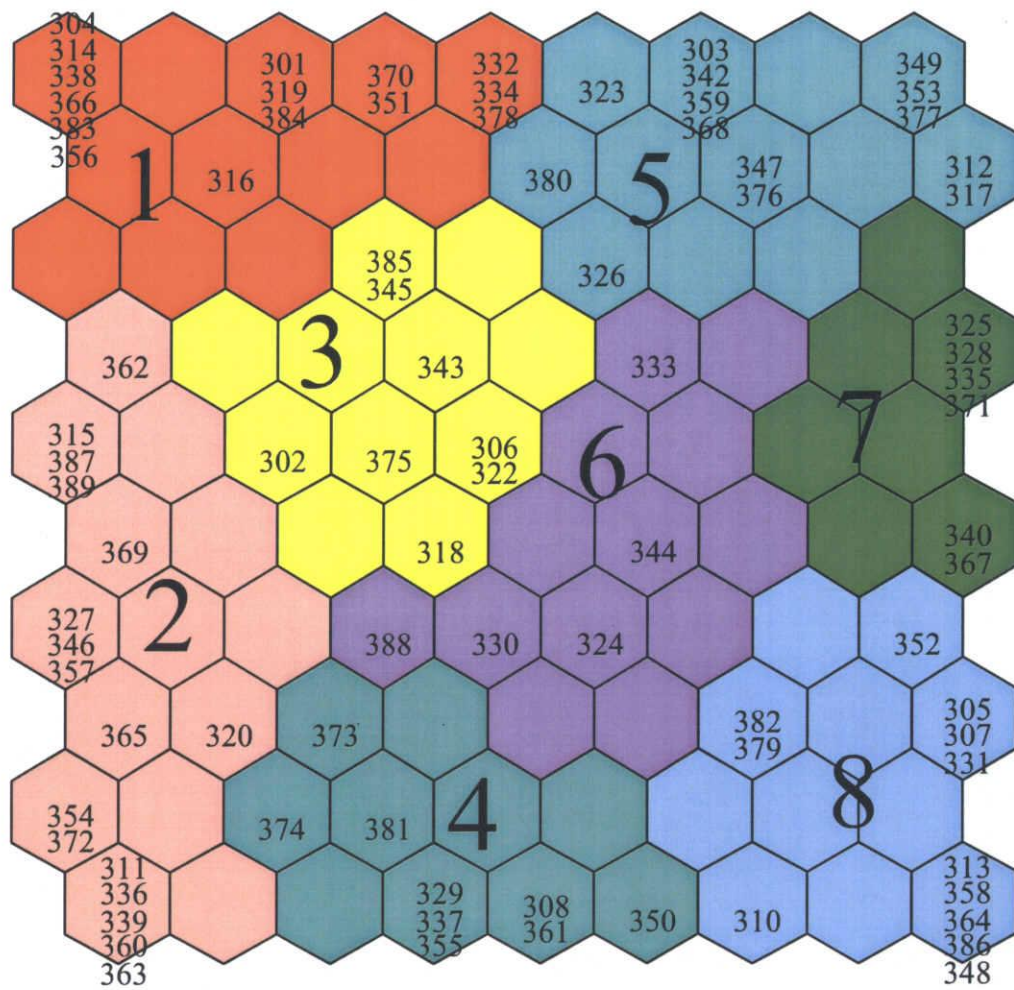


図 5.1: SOM の参照ベクトルに対して k -means 法を適用した結果

表 5.2: 各クラスターの代表の解答 (1/2)

クラスター	クラスターの代表の解答	人手で分けた人数
1	インターネットで何かについて調べる時、たくさんの情報があるが、すべて正しいとは限らないので、正しいか正しくないかは自分で判断しなければならない。	15
2	著作権に関して気をつけなければいけないと思う。レポートなどを作るときにウェブページから文章や写真などを勝手にコピーしてつかうと著作権の侵害になってしまうので文章は自分でアレンジして使い、写真はフリーページからとってくるべきだと思う。	17
3	インターネット上で買い物などをしようとしている時、個人情報を入力しようとする時。時にインターネットカフェなどでは個人情報が流出し安いので出来るだけ、直接書き込まないようにすべきである。	8
4	重要なもの:著作権の侵害 自分のホームページの日記などを更新するとき、個人の Web ページから文書をコピーしてはりつけたりはしないようにしている。その Web ページには、作成者の権利があって、それは国際的に保護されている。よって、勝手に使うことは、法的に問題があり、罪を犯すことになる。その情報をホームページにのせるためには、その事実は変えずに自分の意見などを書くといと思います。	9

表 5.2: 各クラスターの代表の解答 (2/2)

クラスター	クラスターの代表の解答	人手で分けた人数
5	ある事を調べようとした時、インターネットなどを利用して情報を得ようとするならば、その得た情報を容易に信用してはならない。なぜなら、インターネットなどは様々な人々が利用しているため、情報に誤りがあることがあるからだ。	14
6	・著作権の濫用 ・あることをインターネットで調べていて、その調べていることについてのインターネットのページやその一部を自分のインターネットのページなどに無断で載せると著作権の濫用になってしまうので、ちゃんと許可をとらなければいけない。	5
7	インターネットを利用して、自分が調べたい情報を探すときに、サイト内に記載されている情報を無断で公の場に使用すると著作権を侵害してしまうため、そういった情報は、どこから手に入れたのかなどを明記して使用することに気をつけなければならない。	6
8	・使用者のプライバシーの保護が最も重要だと思う。 ・掲示板等の書き込みの際し、他人を中傷するようなことを書くと、名誉毀損になりかねない。	12

表 5.3: 各クラスタに対応したラベル

クラスタ	人手で分けたときの 各クラスタのラベル
1	D
2	B
3	A
4	B
5	D
6	B
7	B
8	A

表 5.4: ラベル B の適応率と適合率

	再現数	再現率	検索数	適応率
平均	28.4	81.1	37.2	78.9
最大	35	100	58	96.4
最小	23	65.7	26	58.6

表 5.5: ラベル D の適応率と適合率

	再現数	再現率	検索数	適応率
平均	23.2	82.9	31.2	75.4
最大	28	100	41	90.9
最小	20	71.4	22	68.3

表 5.6: 各試行回数に対する一致率

試行番号\ラベル	A	B	C	D
1	0	100	100	66.7
2	なし	80	100	100
3	100	100	100	100
4	0	100	0	100
5	100	75	なし	100
6	なし	100	100	100
7	なし	100	100	100
8	0	75	なし	100
9	なし	75	0	100
10	0	100	なし	100

5.2.1 考察

講師は、図 5.1 からクラスタの位置関係と、各クラスタの人数を把握することができ、表 5.2 から各クラスタの内容を知ることができる。例えば、講師は表 5.2 を見てその代表の解答の内容を把握できるが、その代表の解答同士の間関係というのはいはわかりにくい。代表の解答同士の間関係を知るためには、図 5.1 を見ることでわかる。図 5.1 では、クラスタ 1 とクラスタ 8、クラスタ 2 とクラスタ 3 はマップ上の逆の位置にある。そのため、各クラスタの内容が違ふ内容であると推測できる。実際、それらの代表の解答の内容は表 5.3 を見ると、ラベル D とラベル A、ラベル B とラベル A となっており違ふ内容である。また、図 5.1 では、クラスタ 1 とクラスタ 5、クラスタ 2 とクラスタ 4 は隣り合っている。その結果、各クラスタの内容が近い内容であると推測できる。表 5.3 を見ると、クラスタ 1、5 はともにラベル D、クラスタ 2、4 はともにラベル B となっており似た内容である。このように、 k -means のクラスタ数が実際の内容の種類数に比べ多い場合には、同じ内容の解答がいくつかのクラスタに分かれて現れるが、実際には同じ内容のクラスタであることを、比較的容易に把握できる。このように、講師は図 5.1 と表 5.2 を合わせて見ることで、クラスタの主な内容とクラスタの位置関係、各クラスタの人数を知ることができる。以上により、講師は小テストの解答の傾向を把握することができるといえる。

表 5.4, 表 5.5 の結果を見ると, ラベル B, ラベル D 両方とも最大と最小の差が大きく実験結果の分類性能が不安定であることがわかる. しかし, 本研究の目的は講師が小テストの結果のフィードバックをすばやく行うために主要な内容を表示させるものであるので, 小テストの結果の多数ある解答がわかればよい. 表 5.6 を見ると, ラベル B, D いずれも, すべての試行においてその内容を持つ解答が多数を占めるクラスタが存在し, かつその代表の解答として正しくその内容を持つ解答を代表の解答として選択できていることがわかる. 一致率の数値としては最低のもので 66.7% であり一見すると精度が低いように見える. しかし, 全クラスタ数が 8 であり, 指定した内容の解答が多数を占めるクラスタは多くても 5 個であった. $R = 66.7\%$ なら 3 個のクラスタのうち 2 個のクラスタにおいて適切な代表の解答を選んでおり, 全解答の概略をつかむのには十分な精度である.

以上より, 図 5.1 より, クラスタの関係がわかり, 表 5.2 の 8 つあるクラスタから, 主要な内容を把握でき, その概数がわかるようになった.

第6章 まとめ

講師がわかりやすい授業を行うために記述式の小テストを行い，その解答の傾向を把握しすばやく学生にフィードバックを行うことを実現する．本研究では，講師が記述式小テストの解答の傾向を行いやすくするために，以下の点に重点をおいて研究をした．

1. 主要な内容を把握しやすいこと．
2. それらの解答数を把握しやすいこと．

これを実現するために，以下のことが実現できる分類法とその結果の表示方法を提案した．

1. クラスタの関係がわかるように分類結果を2次元で表示する
2. 主要な内容を表示する

結果は，講師が実験結果を見ることで，多くの解答を含む内容についての解答の数とその内容がわかるようになった．

謝辞

本論文は、著者が三重大学大学院工学研究科博士前期過程に在学中に行った研究をまとめたものである。本研究を進めるにあたり、懇切丁寧なご指導とご督励を賜った三重大学林照峯教授，北英彦助教授，高瀬治彦助手に深く感謝いたします。また，日頃熱心に討論して頂いた計算機工学講座の皆様方にお礼申し上げます。

そして，貴重な時間を割いて本研究論文の査読をして頂いた情報処理研究室の鶴岡信治教授に深く感謝いたします。

最後に，本論文をまとめるにあたり，助言，討論，その他お世話になったすべての方々に感謝いたします。

参考文献

- [1] 大井健太郎, 高瀬治彦, 森田直樹, 北英彦, 林照峯: 記述式小テストの解答自動分類のための特徴抽出に関する一考察, 2006 PCカンファレンス論文集, pp.449-452 (2006)
- [2] 藪兼智英, 井上正人, 金明旭, 岡隆光, 前原俊信: 自己組織化マップ (SOM) による文書の分類とキーワードの抽出, 呉大学社会情報学部紀要社会情報学研究, pp. 99-106 (2004)
- [3] 長尾真: 自然言語処理, 岩波書店 (1996)
- [4] 黒橋禎夫, 中村順一, 長尾真: 構文情報を利用した電子ニュース記事のクラスタリングシステムの作成と評価, 電子情報通信学会技術研究報告, 言語理解とコミュニケーション, Vol.98, pp. 15-22 (1998)
- [5] 徳永健伸: 情報検索と言語処理, 東京大学出版会 (1999)
- [6] 工藤拓, 松本裕治: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol.43, No.6, pp. 1834-1842 (2002)
- [7] 元田浩, 津本周作, 山口高平, 沼尾正行: データマイニングの基礎, オーム社 (2006)
- [8] 神寫敏弘: データマイニング分野のクラスタリング手法 (1) クラスタリングを使ってみよう!, 人工知能学会誌, Vol.18, No.1, pp.59-65 (2003)
- [9] Richard O. Duda, Peter E. Hart, David G. Stork: パターン識別, 新技術コミュニケーションズ (2003)
- [10] T. コホネン: 自己組織化マップ, シュプリンガー・フェアラーク東京 (2005)

- [11] Kohonen, T., Kaski, S., Lagus, K., Salojarvi, J., Honkela, J., Paatero, V., and Saarela, A.: Self-organization of a massive document collection, IEEE Trans. Neural Networks, 11-3, pp. 574-585 (2000)
- [12] Juha Vesanto, Johan Himberg, Esa Alhoniemi, and Juha ParHankangas: SOM Toolbox Matlab 5, <http://www.cis.hut.fi/projects/somtoolbox/> (2000)

付録A 実験で用いた解答

この付録では、実験で用いた解答の情報を A.1 にすべて記す。記した情報は、解答を人手で分けたときのラベル、解答に対応する番号、そして、解答の全文である。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (1/17)

ラベル	番号	解答
A	302	個人情報の保護。オークションやオンラインゲーム等で身分証明を行う時に入力するがその時周りの人に見られたり、ウィルス対策のなされていないパソコンによって、情報が他の人に知られてしまうと、それを悪用される危険がある。したがって、情報を入力する際は、周囲の状況や、サイトの信頼性、安全性などに注意を払って慎重に行うべきである。
A	306	個人情報がネット上に漏れないように普段からむやみに個人情報を与えない心がけが大切である。
A	313	自分の Web などに、他人のプライバシーに関わることを載せるときに、相手の許可をしっかりとることが重要だと思う。
A	322	インターネット上で買い物などをしようとしている時、個人情報を入力しようとする。時にインターネットカフェなどでは個人情報が流出し安いので出来るだけ、直接書き込まないようにすべきである。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (2/17)

ラベル	番号	解答
A	344	掲示板などに書き込む時、自分や友達などの情報を不様意に書かないようにするのが重要だと思います。特に他人の情報は本人の合意なしには絶対に書き込まないようにしなければなりません。何故なら個人情報をのせてしまうと悪用される可能性がある上、一度流出した情報は完全に消しざることができないからです。また他人の情報の場合は自信も法で罰せられれしみます。
A	358	・プライバシー ・インターネットは世界中の人が見ることが可能なので、プライバシーを侵害したするようなこと書き込むと、世界中の人に知られてしまうから。
A	364	・使用者のプライバシーの保護が最も重要だと思う。 ・掲示板等の書き込みの際し、他人を中傷するようなことを書くと、名誉毀損になりかねない。
A	370	・個人情報流出 ・インターネット上は様々な情報が流れていて、いつ、誰が見ているのか分からないので、個人の情報を出すときには十分注意する。もし何も考えずにインターネット上に公表すると、その情報を知らない誰かが盗み、悪用されるかもしれない。なので、もしどうしても載せたいときは本当に信用できるところかを判断しなければならない。できる限り公表しないほうがよい。
A	382	プライバシー 例えばインターネット上など不特定多数の人が見る可能性がある場所で他人の情報を許可なく流す事は、法律にふれるだけでなく、被害者がの人権を無視する行為だ。まだまだ日本はそういった法律はゆるい所があるので、しっかりと見直すべきだ。
A	386	最も重要だとおもうものはプライバシー。
B	308	自分の趣味のホームページを作る時に、他人の写っている写真や既存のキャラクターを勝手に使用してはいけない。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (3/17)

ラベル	番号	解答
B	311	人の情報やWebページなどを利用して自らの資料を作成するときに著作権の侵害にならないように気をつけるべき。
B	312	レポートなどつくる時に、他人のホームページからかかって文などをとってくると著作権のしんがいになるのでかかってにとってきてはいけない。もしとりたかったら謝辞をいれないといけない。
B	315	自分が何かの資料やレポートをさくせいすることについて、どこかのサイトから文章を引用したりする時に、その文章の作成者に許可をもらうべき。
B	317	インターネットで音楽などをとってくるときに著作権の事を考えなければいけない。もし歌詞などを無断でとってきた時は著作権侵害で書類送検などもあるから。
B	318	インターネットを使えば容易に様々なデータを入手することができるがどのようなデータに関しても利用する際には著作権が存在していることを忘れてはならない。
B	320	著作権について、私がPCをいじっているとき、画像や文章など著作権にひっかかっていないか心配になります。だから私は他から引用しないでおこうと決めました。
B	325	画像や作文などをコピーまたはダウンロードするときに著作権に気をつける。
B	327	そのホームページの分の引用を行うとき、その文をどこから取ってきたかを明らかにするべきだと思う。
B	328	著作権の侵害について気をつけるべきだと思う。最近では個人のホームページに写真、絵画、イラスト、論文、音楽などを無断で使用することがある。このようなことが当然のように行われているために罪の意識がほとんどないように思われる。今後、小さな頃からしっかり「情報倫理」について教育すべきだと思う。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (4/17)

ラベル	番号	解答
B	329	知的財産権。まずインターネットを通じて世界各地の人々のニュースを見たりできるようになった。でも、それはみんながいつどこで見ているか分からない。しかも、今の技術は進んでいて、それをコピーしようと思えばいつでもコピーできる。そのことについて、ほしかつたら、いつでもコピーすることができるし、その写真な写真が好きだから、自分のホームページなどにものせたくなくなってしまいがちちょっとまってほしい。勝手にコピーして勝手に自分のホームページにのせたら知的財産権の問題がからんでくるのではないだろうか。コピーするだけならいいが、自分のホームページにのせたらそれはやはり違法なので、インターネットなどを使う人は情報倫理の考え方にたってから行動すべきだと思います。
B	335	自分が最も重要であり気をつけたいと思うことは肖像権についてです。人の写真であれば肖像権のことを考えるべきだと思います。自分のホームページに友達の写真などを勝手に載せたら迷惑がかかったり、悪用されかれかねません。自分の写真であるならばホームページに載せるのは自分の勝手だけれども良からぬ人がWeb上にバラまいたり、加工して悪用するといった事態も起こりかねないので控えるべきだと思います。その点で人の写真をホームページ上に載せることは危険で肖像権について深く考え、気をつけなければならないと考えます。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (5/17)

ラベル	番号	解答
B	336	私はインターネットを利用するときは著作権に気をつけなくてはいけないと思う。計算機の授業でICタグについてのスライドを作成した時、Web上の図やグラフをコピーして、貼りつけてしまった。このスライドは授業以外では発表しないが、もし発表すれば違法となってしまう。よってインターネットを利用するときは著作権法に犯さないように気をつけなくてはいけないと思う。
B	337	自分でホームページを作るとき、勝手に人のホームページにある画像や、他人が写っている写真などを貼ると、著作権の侵害になるので、張らない方がいい。
B	339	著作権や肖像権。レポートを書こうとしているときなどにインターネットでみつけたサイト内の文書、図、写真などをかっけてコピーして使ったりすると著作権や肖像権を侵害してしまうので、むやみにコピーしたりしないようにする。
B	346	学校のレポートを提出するために、インターネットから図やグラフ、文章などを引用するときに、これらの著作権に気をつけ、使用に関しての文章や謝辞を記しておくべきであるということ。
B	348	情報倫理に関して気をつけなければいけないことで、自分が最も重要だとおもうのは「著作権」だと思います。例えば自分がある情報についてのレポートを作っているときその情報が書かれた本をまるまる写したら著作権の侵害になります。そうならないためにも、自分の文書で書き、参考にした本はしっかりと書くべきです。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か?」という質問に対するすべての解答 (6/17)

ラベル	番号	解答
B	354	・著作権 ・レポートを作成するような時に資料を他のホームページからもってくる場合に、その資料にも著作権があるから、違法にならないように資料をさがし、うまくレポートに記述しないとイケないから。
B	355	重要なもの：著作権の侵害 自分のホームページの日記などを更新するとき、個人のWeb ページから文書をコピーしてはりつけたりはしないようにしている。そのWeb ページには、作成者の権利があつて、それは国際的に保護されている。よって、勝手に使うことは、法的に問題があり、罪を犯すことになる。その情報をホームページにのせるためには、その事実は変えずに自分の意見などを書くことよいたと思います。
B	357	・著作権の侵害 ・学校のレポートを書く時、他人のサイトの参考資料を使いたいとき勝手に使わずに許可を得てから使うようにする。
B	360	著作権に関して気をつけなければいけないと思う。レポートなどを作るときにウェブページから文章や写真などを勝手にコピーしてつかうと著作権の侵害になってしまうので文章は自分でアレンジして使い、写真はフリーページからとってくるべきだと思う。
B	361	・著作権に関して 例えば自分の好きなアーティストについてホームページを作ろうと思ったときに、音楽や、アーティストの写真などをホームページにのせてよいのかどうか、よく考える必要がある。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (7/17)

ラベル	番号	解答
B	362	自分が何かについての公の発表をする時に、あるサイトにあった文章、図を借用して作った場合、著作権法に違反しないように、参考文献及び参照サイトの欄にサイト名ならびに URL を書いて、敬意を表すべきである。
B	363	インターネットを使って自ら情報を配信するとき、一方的に情報を流してしまうのではなく、プライバシー管理や著作権などに問題がないかよく確認すべきだと思います。インターネットでは文章や発言を取りしめることが難しいので、個人個人がそういった考えをもって行動することが必要だと思います。
B	365	・著作権 ・今世間でさわがれている WINNY を使うとき気をつけなければならない。私は使ったことはないのだが、もしそれを使って映画などを取ってしまうと自分の著作権ではないので、他人の著作権の侵害につながるから絶対にしてはならない。
B	369	レポートなどでインターネットの文章を引用するとき、きちんと謝辞を書くため、その情報源の所をメモ書きにおく。
B	371	インターネットを利用して、自分が調べたい情報を探すときに、サイト内に記載されている情報を無断で公の場を使用すると著作権を侵害してしまうため、そういった情報は、どこから手に入れたのかなどを明記して使用することに気をつけなければならない。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (8/17)

ラベル	番号	解答
B	372	著作権が重要だと思います。自分が大学のレポート宿題がでて、そのレポートを完成させるためにインターネットを利用するとき、勝手にちょうどいい情報などを自分の物のようにつかおうと、著作権の侵害になってしまうからです。インターネット上の情報や画像などを使う時はそれに著作権があるかどうか確認したり、その持ち主に了承を得るべきだと思います。
B	373	・画像を取り込むとき、取り込んだ画像は自分で厳重に管理しなければならない。 ・画像を自分のパソコンに取り込むとき、自分でちゃんと管理しないで、もし、その画像がインターネットに流れると、その人が肖像権及び著作権の侵害とされることがあるから。
B	374	やはり自分は著作権だと思う。現在、インターネットでさまざまな情報が容易に得られるため、ついつい無断で引用しそうになったりする。また最近では自分でホームページをつくったり、ブログを作成したりもできるようになった。ここで、著作権について知らない、無視するといった人が出てきている。このままでは何でもありになってしまう。これを防ぐには、法律をきびしくしたり、もっとCMなどメディアを使って、著作権の大切さをアピールすべきだと思う。
B	381	著作権を守ること。自分たちがインターネットなどでホームページなどを楽しんでいるとき自分の気に入った画像があると保存して、自分のホームページにのせたり、利用したりしまいそうになる。しかし、それは他人の作ったものであるから著作権を侵していることになる。どうすればいいかは、著作者に許可を得て、著作権表示をすればよい。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (9/17)

ラベル	番号	解答
B	385	インターネットの中の情報を好き勝手に自分が周りに広げてはいけない。なぜならインターネット上の情報にも著作権があるため。
B	387	・著作権について。 ・レポートを作成している時、参考にした資料の文章か図を用いたい時に著作権について考えます。どうしてもそれらを用いたい時は、作者にメールで許可をもらったりなど、何らかの方法で許可をとらなければならないと思います。
B	388	・著作権の濫用 ・あることをインターネットで調べていて、その調べていることについてのインターネットのページやその一部を自分のインターネットのページなどに無断で載せると著作権の濫用になってしまうので、ちゃんと許可をとらなければいけない。
B	389	著作権 レポートを書く時、インターネットを使って、調べた内容を直接コピをして、自分の文章にいれなくてははいけません。インターネットで調べた内容は著作権があるから。本当は著作者に申請して、許可をもらわなければなりません。個人使用なら、ちゃんと引用のほうを書かなければなりません。
C	305	掲示板というのは、いろいろな人が見るものであるので、人を中傷するような書き込みはやめるべきである。
C	307	言葉遣いに気をつける事。掲示板やチャットなどを利用する時には、お互いの姿が見えない状態で言葉のやりとりをするので、文字のみで会話をする分、何気ない言葉でも相手が不快に思う可能性がある。だから、できるだけ言葉を慎重に選ぶべきだと思う。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (10/17)

ラベル	番号	解答
C	310	他人を誹謗中傷するような書き込みをしない。チャットや掲示板への書き込みは匿名で書き込めるので誰が書き込んだのか分からない。だからといってそんなことをしていいわけがない。言われた人は傷つくし中高生のトラブルで判例もある。インターネットを使うからにはそういった情報倫理を理解した上での利用を心がけるべきである。
C	331	電子掲示板で自分の顔が見えないからといって、相手の中傷する言葉などを使ってはいけない。
C	340	私は、ネチケットを守ることが重要だと思います。例えば、掲示板などでウソをついてブラウザクラッシャを踏ませたりする人がいますが、あまりに気分が悪いです。また、言葉遣いに気をつけないと、それも他人を傷つけることになります。私はインターネットについて、「有益な情報を得るための手段」として考えているので、余計な争いごとは極力避けたいです。
C	352	チャット、掲示板など不特定多数の人の交流の場において匿名性があるからと、誹謗、中傷で相手の人権を侵害しないように気をつけなければならない。モニタごしだからといってモニタの向こうには生身の人がいることを意識しながら交流するよう心がけるべきである。
C	367	私が一番重要だと思うのはモラルです。匿名性が高いので他人を傷つけたりする人がいますが、現実世界と同じく傷つけられた人の気持ちを考えて、するべきではないと思います。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (11/17)

ラベル	番号	解答
C	380	僕がインターネット利用時に最も注意しないといけないと思うことは「間違っただけ情報をかかない。」ことだと思う。インターネットはすごい数の人が見ているので、間違っただけ情報はすぐに広まってしまい、事件をおこすかもしれない。また地震などの災害時のときに生存者確認の掲示板みたいなところに間違っただけ情報やうその情報をかいてしまうと、結果として誰かが悲しむことになってしまう。だからインターネット上にかきこむときはできるだけ確実な情報を書くべきだと思う。
D	301	何かを調べようとしているとき、ウェブ上に書いてある情報についてすべてを信じてはいけない。中にはその情報にはたまに正しくないことも書かれていたり、それを信じてしまうと誤った判断をしてしまうので、本当にこの情報が自分の知りたいことだったのか、正しい情報なのかをしっかりと確認してウェブ上の情報を利用すべきである。
D	303	書きこんだ人がうその情報を流している可能性もあるので見つけた情報をすぐに信用するのではなく、たくさんのページを参考にしたり、取捨選択したりすること。
D	304	信頼できる情報であるか。レポート等の作成において、資料を収集しなければならない時、1つのWeb ページの情報だけを使うと、それは誤りでないと言い切れないので、複数のWeb ページを利用して、確認することが望ましい。
D	314	インターネットを使って、何かを調べている時には、すべての情報が信頼できるわけではないので、簡単に信じてはいけないということ。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (12/17)

ラベル	番号	解答
D	316	インターネットにある情報は、誰でも発信あるいは受信をすることが可能であるから、すべての情報が正確ではない。よって、それぞれの情報を正しいのかどうかは、自分で判断して、その情報を活用しなければならない。
D	319	私が最も重要だと思うことは、インターネット上の情報をすべて鵜呑みにしないということである。自分にとって大事な決断をする時人の情報を参考にすることはよくあり、また大事なことでもある。しかし、インターネット上の情報がすべて本当だとは限らない。そういう時それを鵜呑みにして後悔しても遅いからきちんと、見極めることが大事だと思う。
D	323	検索等で、情報を得ようとする時に、ただ1つの情報をそのまま信じ込むのではなく、様々な情報を見て、それから正しい情報を得なければならない。なぜなら、1つの情報だけでは、個人個人の解釈の違いから情報が多少ねじ曲がってしまっているかもしれないからである。1つのサイト等の情報だけでなく、いろいろなサイトの情報をよく吟味すべきである。
D	324	何かを調べる時、インターネットを利用する場合、みんなが見てるからといって、正しい情報とはかぎらないというのを頭に置いて調べる必要がある。
D	326	調べものをしている時、情報をすぐ受け入れるのではなく、受け取った情報を自分の中で吟味して間違った情報を受け入れr内容に注意しなければならない。
D	332	インターネットに掲載されている情報がすべて正しいわけではない。インターネット上の掲示板を見てその情報を完全に信じ込んでしまい取りかえしがつかなくなる。対策としては安易にその情報を信用しないこと。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (13/17)

ラベル	番号	解答
D	333	調べものをするときに、いくつものページを見てどれらが本当に正しいのかよく見比べなければならない。
D	334	インターネットを利用して、調べものをするときに、1つの情報をうのみにしてしまい、もしその情報が誤ちがっており、それを知らずに、その情報を人に伝えたら、誤ちがった情報を伝えたことになってしまうから、一つの情報だけでなく様々なところから真偽をはっきりさせていかなければならないと思います。
D	338	自分が調べた情報が正しいかどうかいろんなサイトで調べて自分で判断する。インターネット情報の中には間違っただ情報もあるから。
D	342	インターネットの1つだけのサイトの情報を信じこまない。間違っただ情報の可能性があるため。インターネットのサイトは一般の人が作っているものであつて違っただ情報を故意にせよ故意じゃないにせよ、行われている可能性があり、もつと複数のサイトで調べて総合的に判断する必要がある。
D	345	インターネット上には本当でないことも含む様々な情報が流れているので、調査や情報収集をするときには、どの情報が正確であるかを自分でしっかり見極められなければならない。
D	347	調べ物をするときインターネットは非常に便利だがWebサイトがたくさんあつてその情報の中にはうそも存在するのでWebサイトに書いてあるものを全て信用してしまうと間違っただ情報を得てしまう場合があるのでたつた1つの調べ事でもできるだけ多くのサイトを見てその信憑性を確かめた方がよい。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (14/17)

ラベル	番号	解答
D	349	重要だと思うことは全ての情報を信じてはいけない事だと思います。インターネットでニュースを見たりする時、全ての情報を信じてしまうと多くのまちがった情報まで広く伝わってしまうと思います。だからまず情報を得たらそれが真実かどうか考え、一つの所だけでなく、多くのWebで調べるようにすることが大切だと思います。
D	350	インターネットで調べた情報は、必ずしも正しくないということ。理由は、検索エンジンで調べると、その中には、一般の普通の人のがのせている情報があり、その人が正しい知識を持っているかは、分からないから。
D	351	自分がインターネットで知りたいことや気になることについて調べる時、誰かが書きこんだことをそのままのみにせず、自分で考えたり知り合いの意見を聞いたりし、インターネット上のことは極力参考程度にしなければならない。なぜなら、うたがわずにそれを信じていると、たとえネットがあやまったことを述べていても、それを信じこんでしまい、傾よった智識がていちやくしてしまうからである。また、自分はインターネットで見ただけなのに、あたかも自分が経験したかのようなさっかくにおちいってしまう。
D	353	自分が欲しい情報を得る時に、本当にその情報が正しいのかを判断する事が重要。インターネットでは、多くの情報があり、欲しい情報を手に入れる事は容易だが、間違っただ情報を掲載しているのも珍らしくないので、本当にその情報があっているのかをよく吟味しなければいけない。間違っただ情報を鵜呑みにしてしまうとその時の自分の状況しだいでは大変な事が起きてしまうから。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (15/17)

ラベル	番号	解答
D	359	ある事を調べようとした時、インターネットなどを利用して情報を得ようとするならば、その得た情報を容易に信用してはならない。なぜなら、インターネットなどは様々な人々が利用しているため、情報に誤りがあることがあるからだ。
D	366	インターネットで何かについて調べる時、たくさんの情報があるが、すべて正しいとは限らないので、正しいか正しくないかは自分で判断しなければならない。
D	375	情報の信頼性。大学でレポートを提出する時に、インターネット上のものをそのまま信用するのはあまりよくない。それは情報では、誰でもインターネット上に流出することができるものだから。
D	376	・自分が一番重要な情報は、Web 上にある情報を鵜呑みにしてはいけないことだと思います。全ての情報とまではいきませんが、例えば、サッカー選手の特集のページを閲覧する場合は、交際相手や何か悪い疑惑の噂を素直に信用してはいけないということです。というのは、それは話題が欲しい何者かが故意に作ったデマの可能性があるので。これが真意かどうかは、証拠を載せてあるサイトや、サッカー選手本人が何らかの方法で自ら述べるまでは確証が持てないので、様々な情報を合わせて吟味する必要があります。なので、上記の通り、重要なのはインターネット上の情報を鵜呑みにしてはいけないことだと思います。
D	377	情報を手に入れたい時、インターネットの情報は誰もが書きこめるのでむやみに信じるべきではない。

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (16/17)

ラベル	番号	解答
D	378	<ul style="list-style-type: none"> ・大量の情報の中からさまざまな情報をとりだしより確実な情報を得ること. 決して1つの情報にとらわれない. ・もし自分の知らない単語を知らべるときその単語で調べるだけでなく、その中からさらに自分の知りたい用語で複数検索をするべきだ.
D	383	<p>情報の取捨選択 例えば三重県の歴史について調べたいと思い、インターネットで調べる際にネット上では非公式なサイトや個人が勝手に作ったサイトがたくさんあるので、全ての情報をあてにするのではなく信頼できるサイトで情報を得るべきである.</p>
D	384	<ul style="list-style-type: none"> ・会議など大勢の人に自分の調べた情報を発表、プレゼンする時、聞いている人が知らないことを発表するかもしれないので、インターネットを利用して探した情報を正しいものと正しくないものに見分けることが一番大切だと思う.
E	330	<p>分からないことは自分で調べる。分からないことや、困ったことがでてきとき、インターネットで解決策を調べることができる。このとき質問掲示板で同じことを質問した人がいるかもしれない。もしくはインターネットで検索したらすぐに見つかるかもしれない。調べずに掲示板などで質問すると同じ質問ばかり出てきて、掲示板が見にくくなってしまう。さらには人に嫌われてしまうかもしれない。</p>

表 A.1: 「インターネットを利用するときに、情報倫理に関して気をつけなくてはならないことは何か？」という質問に対するすべての解答 (17/17)

ラベル	番号	解答
E	343	個人情報をインターネット上では使わず、ネチケットを大切に考えインターネットを利用する。
E	356	著作権の問題 オンラインゲームをダウンロードする際、そのファイルが信頼できるか注意したい。同時に信頼できたとしてもウイルスを退治できるようにしておきたい。
E	368	インターネット上の情報を利用するときで、ネット上で得た情報が正しいかどうか。またそれらの情報を利用し、文書などを作ったときの著作権の問題で、情報の利用に関し、自分の作った文書によって、文書を見た人間や、文書の引用元の人間にも不利益を生じさせる可能性があるので、自分の作ったものに対する責任を持てるようにしなければならない。
E	379	「マナーを守ること」だと思う。例えば、匿名の掲示板で他人を誹謗中傷したり、プライバシーを侵害したり、法に触れる情報の掲示や品物の取り引き、無断で知的財産を使用する、不正アクセスをする等はしてはいけない。