

採点ミス発見支援システムの開発 ～部分点と採点記号の認識を用いた採点ミス発見手法～

Development of Support System for Finding Scoring Error - Finding Scoring Error Method Using Recognition of Marking Digit and Sign -

西川 雅清[†] 松尾 賢一[†]
Masakiyo Nishikawa Ken'ichi Matsuo

1. はじめに

人手による答案の採点作業過程では、度々採点ミスが発生する[1]。特に、入試レベルにおける採点ミスの発生は、受験者の合否に影響が及ぶため、大きな社会問題として取り上げられ [2], 入試全体の信頼性を損なうことにもつながる。よって、採点者は採点ミスの発見のために、細心の注意を払いながら採点チェックを繰り返し実施する。しかしながら、この採点チェック過程においても、長時間のチェック作業の従事によって、極度の緊張感、疲労や集中力の低下[3]から新たなチェックミスを生み出す恐れもある。

この現状に対して、採点ミスの原因分析や防止対策の検討、採点作業を支援するシステムの開発が検討されてきた。

特に、マークシート方式の導入は、採点作業の負荷や採点ミス低減に対して効果的であり、一般的に実用化に至っている。しかしながら、このマークシート方式は、複数の解答群から適切な解答を選択する多肢選択式に用いられることが多く、受験者の論理的思考を問う試験に適さないことから、未だに記述式による試験の実施は数多い。この記述式試験での採点作業の支援や採点ミス発生の低減に向けて、画面上に表示された解答欄にペンで記述した解答内容に自動採点を実施するシステム[4]、部分点の認識を容易にする記号を導入して、自動的に合計点を算出するシステム[5]などが開発されているが、すでに採点済みの記述式試験の採点作業の支援や採点ミス発生を低減させる有効なシステムの提案には至っていない。この採点作業支援の実現においては、解答の正当判定、解答に対する採点結果の正当判定、設問ごとの解答に対する配点結果の正当判定など、人間と同等かそれ以上の文字や記号に対する高度な認知機能と推論機能が技術的課題であり、現状では解決が困難な課題といえる。

これに対して、認識機能だけに限定すれば、すでに実用化されている郵便番号読み取り装置には、要素技術としてオフライン手書き文字認識処理[6]が利用されている。つまり、答案中の部分点や採点記号は手書き数字や文字であることから、この要素技術を用いて、採点済みの答案上の採点記号と部分点の認識結果から採点結果の不一致を判定することで採点ミスを採点者に提示できると考えられる。

これまでにオフライン手書き文字認識技術が実用に耐えうる認識精度が得られる理由として、文字パターンを取り巻く背景領域が様々であり、指定枠内に郵便番号を記入させる制約を筆記者に課していることがあげられる。この制

約によって、背景領域と数字パターンの分離を容易にし、数字パターンの重畳や接触を回避することで文字パターンを一文字単位で確実に認識処理に与えることを可能にしている。現在、市販化されている手書き OCR (文字認識装置) においても、指定枠内に書かれた手書き文字を認識対象とすることで実用的な認識精度を実現している[7][8][9][10]。

よって、前述の制約条件を満たさない採点済答案上の採点文字や部分点の認識に、そのままオフライン手書き文字認識技術を適用することができない。これを解決するには、印刷された解答欄や鉛筆で書かれた解答が存在する複雑背景をもつ答案からある程度自由に重ね書きされた手書き採点記号や部分点を正確に分離抽出する前処理が必要となる。

本論文では、採点済答案上の手書き採点記号と部分点を分離抽出する前処理手法を提案し、提案手法で分離抽出した採点記号と部分点に対して従来のオフライン手書き文字認識手法による認識精度を調査する。そして、認識結果に基づいて採点結果と部分点の不一致を発見する「採点ミス発見支援システム」を開発し、開発したシステムの評価結果に基づき、採点ミス発見支援の実現性を明らかにする。

2. 答案・採点ミスの定義

図 1 に解答および採点記号と部分点を模倣的に記入した答案用紙 (高専入試解答用紙) 例の一部を示す。

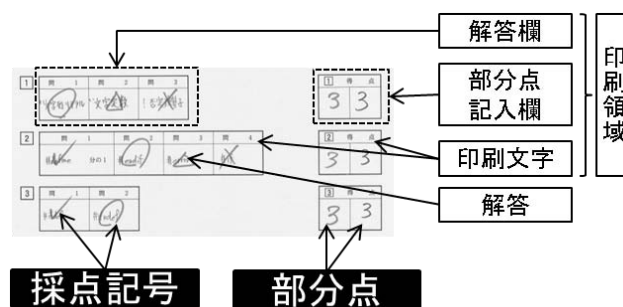


図 1 答案用紙例

図 1 の答案には、解答欄、部分点記入欄、印刷文字からなる「印刷領域」、解答欄内に受験者が記入した「解答」が存在する。採点時には、採点者が解答を読み、正誤を判定し、「採点記号」を解答付近に重ね書きするとともに、採点記号に割り当てた配点に基づいて部分点記入欄内に「部分点」を手書きする。高専入試の採点作業は、採点記号の見誤りや見落としなどの採点ミスの発生を防ぐために、採点者に採点記号同士を接触させず、部分点は枠内に記入する制約を与えている。このように筆記者に制約を与えて

[†] 奈良工業高等専門学校 専攻科 電子情報工学専攻
National Institute of Technology, Nara College
Faculty of Advanced Engineering, Advanced Course of
Electronics and Information Engineering

書かれた文字に対する認識を「制約手書き文字認識」[8]と呼んでいる。

これに対して、一般的な答案では、採点記号や部分点の記入位置に制約を与えず採点記号や部分点が答案に記入されることがある。このような文字に対する認識を「自由手書き文字認識」[8]と呼ぶ。この自由手書き文字認識は、制約手書き文字認識と比較して文字認識率が著しく低下する傾向にある。この自由手書き文字認識の認識率低下の要因として、文字の存在位置の確定に加え、一文字単位での文字切出しが困難であることがあげられる。これは、一般的な答案に見られる重畳や接触した採点記号や部分点同士を1文字単位に良好に分離する文字切出しの問題に帰着する。この文字切出しの問題についても、文字認識の重要な解決すべき技術的課題であるが、本論文においては現段階で取り扱わないものとし、対象とする答案においては、制約手書き文字認識技術を応用した採点ミス発見支援として、採点記号や部分点文字同士の重畳や接触がない答案を対象とする。

次に、採点ミスの事例について述べる。表1に示すように採点手順上でⅠからⅤの採点ミスの発生が考えられる。

表1 採点手順上での採点ミス

採点手順 ↓	I. 正答と誤答の判断を間違える「正誤判定誤り」
	II. 採点漏れなどの「記入誤り」
	III. 採点記号と部分点の「対応付け誤り」
	IV. 部分点を合計する際に起こる「集計誤り」
	V. その他予期せぬ採点ミス

Ⅰ、Ⅱに対する採点ミス発見支援システムの実現には、受験者が記述した手書きの解答文字に対する非常に高精度な認知機能と推論機能の実装が必要であり、現状では解決困難な研究課題といえる。

本論文では、これらの採点ミスの中で現状の要素技術である制約手書き文字認識技術を利用可能であるⅢの「対応付け誤り」に限定した採点ミス発見支援システムの実現を目指す。

図2に、Ⅲの対応付け誤りの採点ミス答案例を示す。採点者は答案用紙に記入した採点記号を基にして計算した部分点を部分点記入欄に記入する。図2の答案では、「○」に1点、「／」に0点が配点されている。

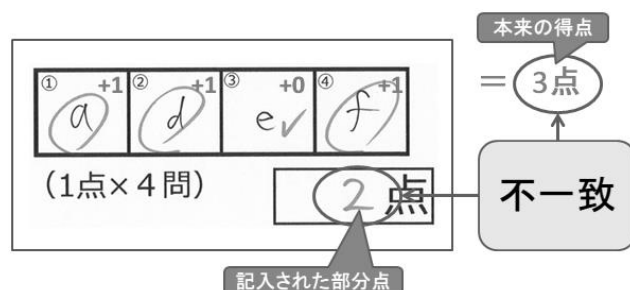


図2 対応付け誤り答案例

したがって、図2では「○」が3つ、「／」が1つ存在することから、本来の得点は3点となる。ところが、部分点記入欄には採点者によって、2点が記入されているのが

わかる。このように、採点記号から算出された本来の得点と部分点の不一致が発生しているミスが対応付け誤りである。

3. 採点ミス発見支援システムの概要

開発する「採点ミス発見支援システム」の概要を図3に示す。本システムは、採点者によって採点作業と採点ミスの確認作業が終了した答案に対して、採点記号と部分点の認識結果の不一致があれば、採点ミスである対応付け誤りの可能性が高い箇所を採点者に提示する。実際の作業では、本システムの出力によって、採点者が対応付け誤りの可能性が高い箇所では採点ミスを確認できたとき、採点の再チェックを実施し、採点ミスが確認できないとき注意喚起となる。図3に示すように採点ミス発見支援システムは、「抽出部」、「部分点認識部」、「採点記号認識部」、「判定部」の4つの処理部から構成される。

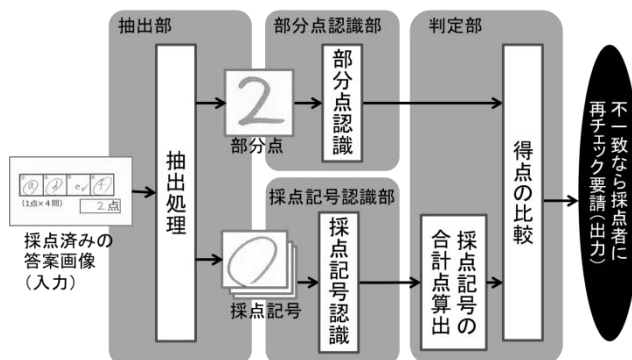


図3 採点ミス発見支援システムの概要

採点ミス確認作業を終えた答案画像を入力として、「抽出部」では、印刷された問題文、解答欄、手書きで記入された解答、採点記号、部分点が重畳する答案画像から、採点記号と部分点パターンだけを分離抽出する。部分点パターンは「部分点認識部」で数字として認識し、得点を算出する。採点記号パターンは「採点記号認識部」で採点記号として認識し、採点記号に対応付けられた得点を用いて、採点記号の合計点を算出する。

最後に、「判定部」において、部分点の得点と採点記号の合計点の整合性をチェックし、整合性が不一致であれば、採点ミスの可能性が高い箇所を採点者に提示し、再チェックを要請する処理の流れとなっている。これ以降に各処理部の詳細について述べる。

3.1 抽出部

図4に抽出部の処理手順を示す。抽出部は、印刷文字、解答欄の枠線、枠線内に記入された解答が存在する複雑背景上に、重ね書きされた採点記号と部分点記入欄に存在する部分点が混在するカラー答案画像から、採点記号と部分点の領域を採点パターンとして抽出する処理部である。採点者によって目視で採点ミス確認作業を終えた答案用紙をスキャナで電子化し、RGB形式のカラー答案画像を抽出部の入力に与える。

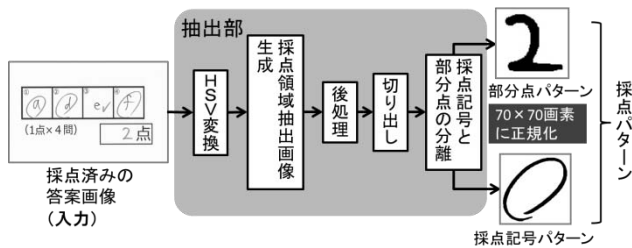


図4 抽出部の処理手順

まず、色情報の処理を容易にするために、カラー答案画像のRGB値から、直感的に色情報を取り扱いやすい、色相H、彩度S、明度Vの3つの尺度をもつHSV値を算出[11]する。

採点作業で採点者は、赤ペンを用いることが一般的であることから、カラー答案画像上の赤ペン筆跡上のHSV値域を事前に取得し、入力したカラー答案画像から赤ペン筆跡のHSV値域をもつ画素を「採点領域」としてカラー答案画像から抽出する。表2に予備調査で得られた採点領域と背景領域のHSV値域を示す[12]。

表2 採点領域と背景領域のHSV値域

	色相H	彩度S	明度V
採点領域	0～30, 161～180	41～255	81～255
背景領域	不定	0～69	不定

採点領域とは別に、紙面、印刷領域、解答から構成される「背景領域」の画素における彩度Sは、予備調査で0～69の値域であった。表2の採点領域と背景領域の彩度Sの値域から、背景領域は無彩色、採点領域を有彩色と捉え、領域分割が可能といえる。さらに、採点領域の色相Hは、赤色の色相を示していることがわかる。よって、カラー答案画像の各画素に対して、採点領域の色相Hと彩度Sの値域をもつ画素を画素値1、それ以外を画素値0とした2値画像である「採点領域抽出画像」を生成する。

複雑背景と採点領域の重畳による混色の影響によって、HSV値が変動することから、採点領域抽出画像内の採点領域のストロークの一部に欠損や断線が発生する。

これに対して、抽出処理内の「後処理」で膨張収縮処理によって採点領域のストロークの欠損と断線を補完する。

後処理後の採点領域抽出画像に8連結のラベリングを施し、得られた同一ラベルの連結成分を1つの採点パターンとして切り出す。この切り出しにおいて、連結成分の面積Aに対して、 $A \leq T$ となる連結成分を採点領域抽出画像から削除し、それ以外の連結成分を「採点パターン」として切り出す。連結成分を除去するしきい値は、 $T=50$ に設定した。

最後に切り出された採点パターンは、縦横70×70画素の大きさに正規化して出力する。

3.2 認識部

図5に認識部の処理手順を示す。認識部は、採点領域抽出画像中の採点パターン1文字を入力とし、出力として採点パターンの認識結果を出力する処理部である。

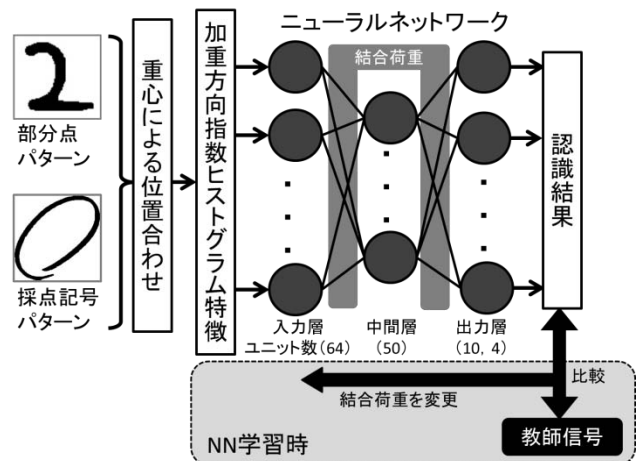


図5 認識部の処理手順

入力された採点パターンに対して、採点パターンの重心位置に基づき位置合わせする。採点パターンの形状特徴を顕著にするために、特徴抽出手法に手書き文字認識に有効とされる加重方向指数ヒストグラム特徴[13][14]（64次元）、識別器として、入力層64個、中間層50個、出力層10個もしくは4個の3層のユニット数で構成した階層型ニューラルネットワーク（以後NNとする）を用いる。このNNは、部分点と採点記号パターンで個別に構築する。出力層の各ユニットに認識対象である採点パターンのカテゴリが対応付けられる。よって、部分点認識用NNでは、採点パターンの認識結果のカテゴリが、「0」から「9」の10種類であるため出力層ユニット数は10個となる。また、採点記号認識用NNでは、認識結果のカテゴリが「○」、「△」、「×」、「/」の4種類であるため、出力層ユニット数は4個となる。NNの学習には、バックプロパゲーション（誤差逆伝播法）を用いる。

このNNにある採点パターン特徴量を入力層に与え、出力結果と教師信号を比較し、入力した採点パターンに対応した正しいカテゴリ名が出力されるようにユニットの結合荷重を変更させる。実際の採点パターンの認識では、学習済みのNNの入力層に1文字単位で切り出した採点パターンの特徴量を入力し、出力されるカテゴリ名が認識結果となる。

3.3 判定部

判定部の処理の概要を図6に示す。この判定部では、図6に示すように認識部で得られた部分点および採点記号パターンに対する各々の認識結果間において、採点結果の整合性を判定する処理部である。

図6において、答案画像の採点パターンに対して、抽出部、認識部を経て、採点パターンの認識結果である部分点と採点記号のカテゴリ名が判定部の入力として与えられる。

各採点記号カテゴリ名は、配点表に基づき点数化され、大問ごとに点数の合計点が算出される。最後に、部分点カテゴリ名と合計点を比較し、整合性を判定する。図6を例に具体的な処理について述べる。入力した答案画像において、抽出部で最初の大問から採点記号が3パターン抽出されたとき、各採点記号パターンが認識部に入力として与え

られ、認識結果である採点記号カテゴリ名が出力される。この採点記号カテゴリ名から配点表によって、‘○’に2点、‘△’に1点、‘×’と‘／’に0点を配点する。そして、配点の合計点として、2点+1点+0点=3点が算出される。この合計点3と部分点欄内に書かれた手書きの部分点を認識したカテゴリ名‘3’を比較する。

このとき、両者が不一致であるとき、採点ミスが発生しているとして、答案画像上でミス発生箇所を採点者に提示する。

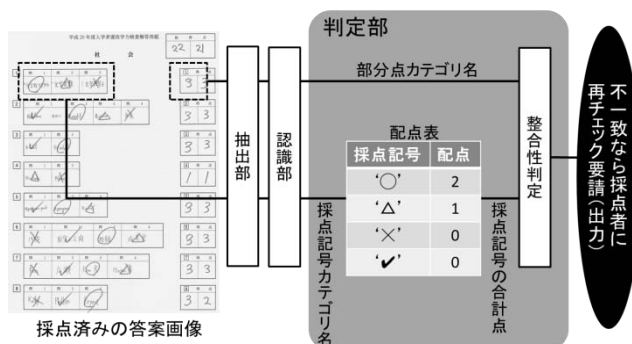


図6 判定部の処理の概要

4. 抽出・認識精度評価実験

ここでは、採点ミス発見支援システムを構成する3つの処理部の内、抽出部、認識部の精度評価を実験によって調査する。さらに、開発した採点ミス発見支援システム全体の性能評価実験によって、採点ミスの提示精度を定量化することで、採点支援の実現性を明らかにする。これ以降で各実験の詳細について述べる。

4.1 実験環境

評価実験で用いる模擬答案用紙を図7に示す。この模擬答案は、高専入試で用いられた解答用紙に、実験協力者によって採点記号と部分点を赤ペンで記入したものである。

図7の模擬答案用紙では、8つの大問ごとに、解答欄が分割された書式であり、各大問に対して2つから4つの解答欄が存在する。この各解答欄に採点記号が実験協力者によって赤ペンで手書きされている。次に、模擬答案用紙内の右端に、大問ごとに部分点記入欄があり、各大問の部分点が実験協力者によって赤ペンで手書きされている。

手書きされた採点記号は、‘○’、‘△’、‘×’、‘／’の計4種類、部分点は‘0’から‘9’の1桁の数字に限定する。また、手書きの場所については、解答欄や部分点記入欄の枠線との接触、重量を許容しながらも、記号や数字の中心位置が欄外にならないよう記入されている。

この模擬答案用紙をイメージスキャナ（brother製 justio MFC-6490CN）を用いて解像度300dpiで電子化して模擬答案画像を作成する。

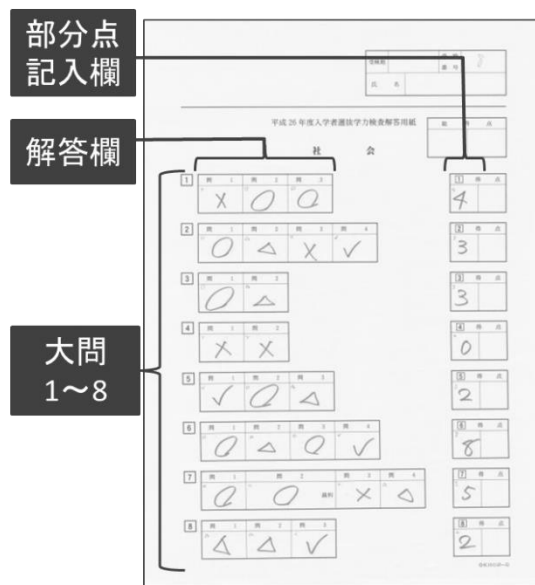


図7 模擬答案用紙例

4.2 採点パターン抽出実験

抽出部における模擬答案画像に対する採点パターン抽出精度を調査する[12]。

4.1の実験環境に基づき、各解答欄に鉛筆で模擬的に解答を記入した模擬答案用紙に対して、赤ペンで採点記号と部分点を手書きする。1枚あたり35個の採点パターンが存在する模擬答案用紙を5枚作成する。5枚の模擬答案用紙に手書きされた全175文字の採点パターンに対して、抽出部によって抽出された採点パターンを目視で確認する。

このとき、抽出された採点パターンに断線や欠損がないとき「抽出成功」と判定する。

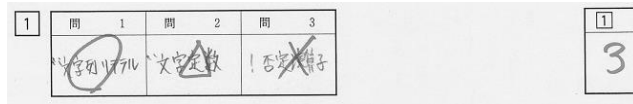
採点パターン抽出実験結果を表3に示す。

表3 抽出実験結果

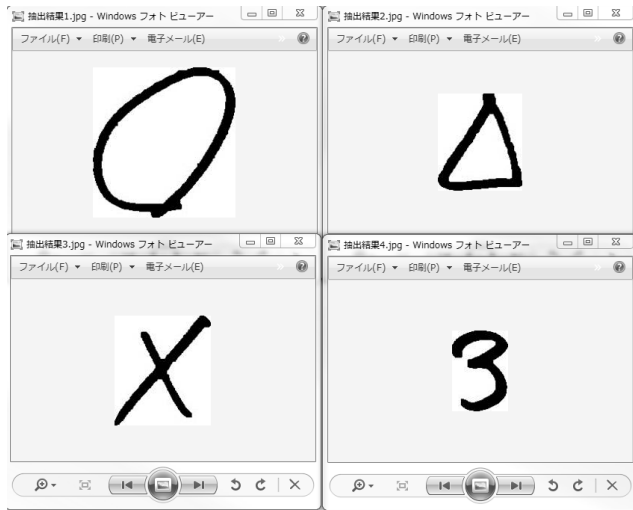
採点パターン数[個]	抽出成功数[個]	抽出率[%]
175	175	100

表3の実験結果から、模擬答案上の全175個の採点パターン全てが抽出成功であり、採点パターン抽出率は100%となった。この抽出実験結果から、表2に示した採点領域のHS値域によって、図8(a)の模擬答案画像から同図(b)のように採点パターンが抽出されることを確認した。採点領域抽出画像中での採点パターンのストロークの一部の欠損や断線に対しては、補完処理によって採点パターンのストロークが修復されていることを確認した。

本実験において、複雑背景である答案に重ね書きした採点パターンに対して表3に示した100%の抽出率が得られた要因は、スキャナによる答案の電子化であった。スキャナで答案を電子化することで、カメラ入力と異なり外乱の影響を受けない電子化が可能であり、結果的に採点領域のHS値域が一定値に収束していることから、簡便な提案手法で高い抽出率が得られたといえる。



(a) 入力した模擬答案の一部



(b) 抽出した採点パターン

図8 抽出処理結果

4.3 採点パターン認識実験

認識部における手書き採点パターンの認識精度を調査する。ここでは、まず採点パターン重ね書きおよび採点パターン同士の重畳の影響を排除した認識精度を調査するため、解答欄に採点記号1つを記入した模擬答案を作成する。

2人の採点協力者A、Bによって、1人あたり10枚、計20枚の模擬答案を作成する。認識対象である採点パターンは1人あたり部分点80個と採点記号250個である。

4.3.1 汎用辞書による認識実験

採点パターン認識に用いる辞書は、部分点と採点記号で個別に作成する。辞書作成には、MNIST手書き数字データベース[15]から入手した1000個の数字パターン（‘0’から‘9’各100個）を利用する。一方、採点記号の辞書作成では、既存のデータベース等が存在しないため、独自にデータを収集する。収集したデータは、模擬答案採点者とは異なる5人の筆記者によって記入された採点記号パターン2000個（‘○’、‘△’、‘×’、‘／’各500個）である。これらの部分点および採点記号をNNに学習させた辞書を「汎用辞書」とする。不特定の採点者の採点パターンの認識には、この汎用辞書を用いる。汎用辞書作成時のNNの学習パラメータを表4に示す。

表4 汎用辞書作成時の学習パラメータ

	部分点	採点記号
安定化係数	0.1	0.1
学習係数	0.1	0.1
学習回数	15万	5万

NN学習時の実行画面例を図9に示す。

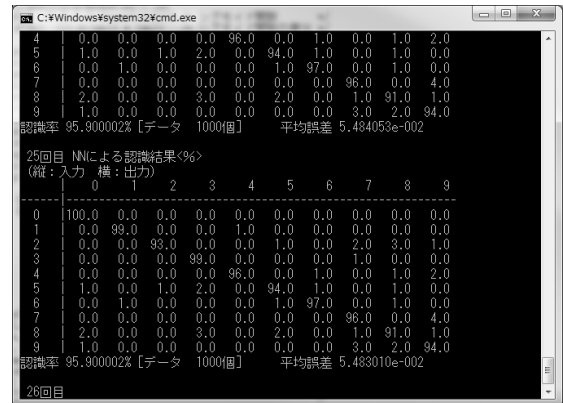


図9 NN学習時の実行画面例

学習時には、学習回数ごとに各カテゴリの認識率と平均認識率、平均誤差が表示され、指定学習回数に達するとNNの結合荷重を辞書として保存し、終了する。

汎用辞書を用いた部分点の部分点認識率を表5に示す。表5の認識率において、辞書作成に用いたパターンを学習パターン、作成に用いないパターンを未知パターンとする[16]。

表5 汎用辞書での部分点認識率

採点者	認識率[%]	
	学習パターン	未知パターン
A	97.6	90
B		77.5
平均	97.6	83.8

表5に示す汎用辞書による部分点の認識実験の結果として、学習パターンで97.6%、未知パターンで平均83.8%の認識率が得られた。

続いて、汎用辞書を用いた採点記号の認識率を表6に示す。

表6 汎用辞書での採点記号認識率

採点者	認識率[%]	
	学習パターン	未知パターン
A	100.0	100.0
B		97.2
平均	100.0	98.6

表6に示す汎用辞書による採点記号の認識実験の結果として、学習パターンで100.0%、未知パターンで平均98.6%の認識率が得られた。

表5と6の結果から、部分点、採点記号ともに学習パターンに対して、未知パターンの認識率が低下し、従来の文字認識手法と同様の傾向[16]を示した。また、部分点と採点記号では、認識カテゴリ数が部分点の方が多いことから両者において部分点の認識は採点記号よりも精度が低くなるのがわかる。

4.3.2 個人辞書による認識実験

汎用辞書は、不特定の採点者の採点パターンが認識対象であった。実際の採点作業においては、1人の採点者が一

定枚数の答案を採点することが多い。このような状況下では、採点者が既知であることから汎用辞書でなく、個人辞書を用いることが可能である。そこで、個人辞書を用いた認識実験を実施する。

まず、採点者によって書かれた部分点および採点記号をNNに学習させた「個人辞書」を作成する。個人辞書の作成には、同一の採点者が筆記した部分点パターン100個（‘0’から‘9’各10個）、採点記号パターン40個（‘○’，‘△’，‘×’，‘／’各10個）を用いる。個人辞書作成時のNNの学習パラメータを表7に示す。

表7 個人辞書作成時の学習パラメータ

	部分点	採点記号
安定化係数	0.1	0.1
学習係数	0.1	0.1
学習回数	10万	5万

個人辞書を用いたときの部分点認識率を表8に示す。

表8 個人辞書での部分点認識率

採点者	認識率[%]	
	学習パターン	未知パターン
A	100.0	86.3
B	100.0	90.0
平均	100.0	88.2

表8に示す個人辞書による部分点の認識実験の結果として、学習パターンで全て100.0%、未知パターンで平均88.2%の認識率が得られた。

続いて、個人辞書を用いたときの採点記号の認識率を表9に示す。

表9 個人辞書での採点記号認識率

採点者	認識率[%]	
	学習パターン	未知パターン
A	100.0	97.2
B	100.0	98.8
平均	100.0	98.0

表9に示す個人辞書による採点記号の認識実験の結果として、学習パターンで全て100.0%、未知パターンで平均認識率98.0%の認識率が得られた。

採点記号においては、表6の汎用辞書と表9の個人辞書の認識結果に差異は見られなかったが、部分点においては、表5と表8の認識結果から汎用辞書よりも個人辞書の認識率の方が向上していることがわかる。この結果から採点記号はパターンの形状が単純であり、筆記者間でのパターン変動が小さいことが考えられる。

また、部分点認識においては、表5より汎用辞書では学習パターン認識率が97.6%、個人辞書では表8より100.0%に向上した。同様に、未知パターンにおいても、個人辞書導入で平均認識率が83.8%から88.2%となり4.4%向上した。これにより、認識部における個人辞書導入は、特に部分点での認識に有効に機能することが明らかとなった。

4.4 システム性能評価

開発した採点ミス発見支援システムの性能評価として、模擬答案を用いた採点ミス提示精度を評価実験によって明らかにする。

4.4.1 性能評価方法

認識部に用いた汎用辞書と個人辞書の2種類に分けてシステム性能評価を実施する。システムに入力する画像は、認識実験で用いた解答欄に採点記号のみを記入した模擬答案を用いる。大問の数は合計160個（模擬答案1枚あたり大問8個×10枚×2人）であり、採点ミスは答案内に合計38ヶ所存在する。答案用紙の書式は既知とする。

ここで、システムの性能評価の指標として、項目 α 、 β 、 γ 、 σ を以下のように定義する。

α ：大問の数

β ：模擬答案に含まれる採点ミスの数（正解群）

γ ：システムが提示した箇所数（出力）

δ ：システムが提示した採点ミスの数（正しい提示）

この各項目の関係は図10のように表され、項目 β 、 γ 、 σ によって、再現率R（Recall）と適合率P（Precision）を、

$$\text{再現率 } R = \frac{\delta}{\beta} \times 100 \quad (1)$$

$$\text{適合率 } P = \frac{\delta}{\gamma} \times 100 \quad (2)$$

によって算出する。

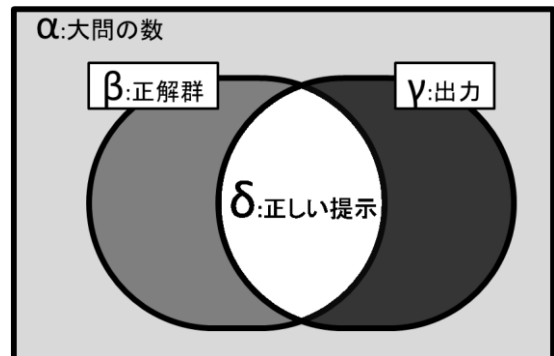


図10 項目 α から δ の関係

再現率Rは、システムが提示するべき正解群の中で採点ミスの正しい提示を表す網羅性の指標である。また、適合率Pは、システムが提示した箇所内に正解群が含まれている正確性の指標となる。

4.4.2 汎用辞書を用いたシステムの性能評価

システム動作時の採点ミス提示結果の画面例を図11に示す。実行結果として、入力された全答案の各大問に対する採点ミスの有無を表示する。この採点ミス提示結果と実際の答案上での採点ミスの有無を比較し、システムの性能を評価する。

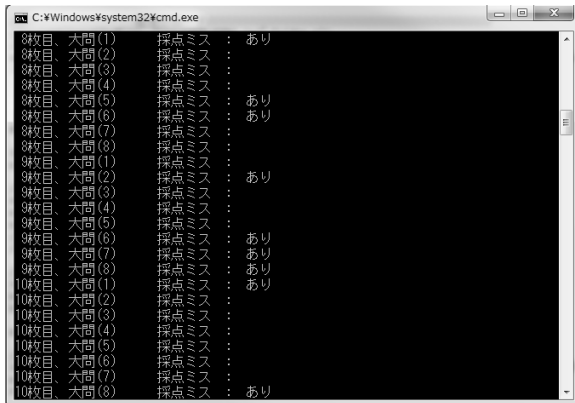


図 11 採点ミス提示結果の画面例

汎用辞書を用いたシステムでの性能評価の指標である各項目数を表 10 に示す。

表 10 汎用辞書での各項目数

採点者\項目	α	β	γ	δ
A	80	18	23	18
B	80	20	34	19

表 10 において、採点者 A、B 共の全模擬答案中の大問の数 α は、80 問（8 問×10 枚）である。

採点者 A において、大問の数 80 の中に、18 箇所の採点ミスが存在する。これに対して、システムの出力は、23 箇所の採点ミスがあり、その中で 18 箇所が正しい提示であった。つまり、模擬答案に含まれる採点ミス 18 のうち、システムが提示した採点ミスは 18 箇所であり、全ての採点ミスを提示したことがわかる。

同様に、採点者 B において、20 箇所の採点ミスが存在し、システムの出力は、34 箇所の採点ミスを提示し、そのうち 19 箇所が正しい提示であった。模擬答案に含まれる採点ミス 20 から、システムが提示した採点ミスは 19 箇所であり、1 問の採点ミスを見逃したことがわかる。

ここで、本来採点ミスでない箇所を採点ミスであると提示した数は、システムが提示した箇所の数 γ からシステムが提示した採点ミスの数 δ を引いた数（ $\gamma - \delta$ ）であり、採点者 A で 5 箇所、採点者 B で 15 箇所であった。

次に、表 10 の評価結果から算出した再現率と適合率を表 11 に示す。

表 11 再現率・適合率（汎用辞書）

採点者	再現率 R[%]	適合率 P[%]
A	100	78
B	95	56
平均	97.5	67.0

採点者 A では再現率 R は 100% を示し、採点ミスの見逃しがあった採点者 B で再現率 R は 95% であった。また、本来採点ミスでない箇所を採点ミスであると提示した数は、採点者 B の 15 箇所に比べ、採点者 A では 5 箇所と少なかったことから、採点者 A の適合率 P の方が採点者 B より 22% 高くなっていることがわかる。

4.4.3 個人辞書を用いたシステムの性能評価

個人辞書を用いたシステムでの性能評価の指標である各項目数を表 12 に示す。

表 12 個人辞書での各項目数

採点者\項目	α	β	γ	δ
A	80	18	30	18
B	80	20	32	19

また、汎用辞書同様に表 12 の項目数に基づき算出した再現率 R と適合率 P を表 13 に示す。

表 13 再現率・適合率（個人辞書）

採点者	再現率 R[%]	適合率 P[%]
A	100	60
B	95	59
平均	97.5	59.5

採点者 A において再現率 R は 100%、採点ミス見逃しがあった採点者 B で再現率 R は 95% であった。また、適合率 P は、汎用辞書を用いた実験結果に比べ、採点者 A では 78% から 60% となり 18% 低下し、採点者 B では 56% から 59% となり 3% 向上した。

システムは、採点者が採点チェックの段階で発見できなかった採点ミスを提示することで採点ミス確認作業を支援するため、システムが提示すべき正解群である採点ミスの見逃しを減らし、再現率 R を向上させることが重要となる。

性能評価の結果、システムが提示しなかった採点ミスは、汎用辞書、個人辞書ともに、採点者 A では 0 問、採点者 B では 1 問のみであり、いずれの条件においても再現率 R が 95% 以上であった。

以上の結果から、開発した採点ミス発見支援システムによる採点ミス発見支援の実現性が明らかになった。

5. おわりに

本論文で採点済み答案上の部分点と採点記号の分離抽出手法を提案した。そして、従来の手書き文字認識手法による部分点と採点記号の認識精度を明らかにすると共に、認識結果に基づいた採点結果と部分点の不一致の発見を可能にする「採点ミス発見支援システム」を開発し、開発したシステムの評価結果に基づき、採点ミス発見支援の実現性を明らかにした。

まず、採点パターンの抽出実験では、模擬答案上の全 175 個の採点パターンに対して、100% の抽出率が得られた。

また、この抽出実験で、印刷枠、印刷文字、枠線、解答が存在する答案に重ね書きされた採点記号を分離抽出可能であることを示した。実際の試験答案では、紙質、筆記具の種類、ペン色が様々に変化するため、答案の色情報に対するクラスタリングを導入し、動的に採点領域と背景領域を分割する処理を導入させてロバスト性を向上させる必要がある。

次に、採点パターンの認識実験では、抽出した採点パターンである部分点と採点記号に対する認識精度を示した。

認識率自体は、これまでの制約手書き文字認識と同水準の結果が得られた。部分点の認識では、汎用辞書よりも個人辞書を用いた方の認識率が高くなった。一方、採点記号

の認識では、両辞書間に認識率の差は見られなかった。これらの結果が得られた理由として、採点記号は、個人、他者間共にパターン同士の変動が小さく、逆に部分点のそれは変動が大きいこと、認識文字種であるカテゴリ数が採点記号の 4 文字に比べ、部分点は 10 文字と多いことがあげられる。これらの理由により、認識率および再現率 R の向上の点から、可能な限りシステム内に各採点者の個人辞書を構築することが望ましいといえる。

また、開発したシステムの性能評価において、汎用辞書を用いたときの再現率 R は平均 97.5%、適合率 P は平均 67.0%、個人辞書を用いたときの再現率 R は平均 97.5%、適合率 P は平均 59.5%であった。よって、いずれの採点者や辞書を用いても 95%以上の再現率 R であったことから、採点ミス発見支援システムの実現性を示したといえる。

この再現率 R は、システムに導入した手書き文字認識処理ルーチンの認識精度に強く依存する。現状 100%の認識率が得られるオフライン手書き文字認識手法は存在せず、人間自身も文字を 100%認識できないことから、全ての文字や筆記者の増加に対して、100%の再現率 R を維持するシステムの実現は困難といえる。したがって、完全な採点ミスの発見よりも、採点ミス発見を支援するシステムの開発を目指す方が、システムの実現性が高いといえる。

さらに、実用性の観点から、ある一定の再現率 R を備えた採点ミス発見支援システムに対して、実際に採点者に対してミスの箇所を提示し、より良い支援であると感じる適合率 P を官能実験で明らかにする必要がある。さらに、再現率 R と官能実験で得られた適合率 P の両値から、

$$F = (2 \times P \times R) / (R + P) \quad (3)$$

の式によって F 値を算出することで、システムの実用性の基準を定量的に示すことが必要である。

今後の課題について述べる。一つ目は、2 つの採点記号同士が接触あるいは重ね書きされた重畳採点記号が存在する答案に対して、採点ミス発見支援を可能にするシステムの改良があげられる。

現在、筆者らは重畳採点記号に対する分離手法[17]は、すでに提案済みだが、分離後の 2 つのパターンが本来分離されるべき採点記号であるか否かの判定が困難であった。

そこで、開発した採点ミス発見支援システムの一部である採点記号認識部を分離手法の後処理に導入することで、分離後の 2 つのパターンと辞書間の類似度値から採点記号として判定可能であると考えられる。今後、この認識部を後処理に追加した分離手法を現在の採点ミス発見支援システムの前処理に導入し、重畳採点記号が存在する答案に採点ミス発見支援の適用範囲を拡大できるかを検証する。

二つ目は、大量の評価用答案データの収集である。現段階で実験協力者 2 名のデータだけであり、模擬答案枚数、学習パターン数、未知パターン数、辞書パターン作成のいずれにおいても、再現率 R や適合率 P の精度評価の観点から十分な採点記号や部分点パターン数を収集できていると言えない。今後は、実験協力者数を増員し、データサンプル数を大幅に増やすことで、精度評価の信頼性および NN 学習性能の向上を目指す。

謝辞

本研究は、JSPS 科研費 15K00455 の助成を受けたものである。

参考文献

- [1] 学力検査における採点・集計ミス再発防止検討委員会, “学力検査における採点・集計ミス再発防止検討委員会まとめ”, (2009) <<http://www.hyogo-c.ed.jp/~koko-bo/saihatubousi/matome.pdf>>, (参照 2013-12-16) .
- [2] 大阪府, “入学選抜における採点方法の改善について”, (2013). < <http://www.pref.osaka.lg.jp/kotogakko/gakuji-g3/kaizen.html>> (参照 2013-12-16)
- [3] 津川 知久, “入学選抜検査採点ミスに関する要請書”, (2009) <http://www.hyogo-kokyoso.com/_src/sc552/2009_04_28_no01.pdf>, (参照 2013-12-16) .
- [4] 曾谷 俊男, “ペン入力によるドリル答案作成/採点システムプロトタイプ”, 情報処理学会研究報告. コンピュータと教育研究会報告, 95,111,pp.49-54,(1995).
- [5] 宮原 隆行, “記述式試験採点支援システムにおける紙メディアに対する入力ミス訂正方式の一提案”, 全国大会講演論文集 第 70 回平成 20 年(4), pp.85-86, (2008).
- [6] 堤田 敏夫, 城戸 賛, 太田 一浩, 木村 文隆, 岩田 彰, “文字認識研究の新たな展開に向けて: 郵便番号データにみる手書き数字認識の現状”, 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, 96(598), pp.73-78, (1997).
- [7] 梅田 三千雄, “手書き漢字認識手法の個性的なサーベイ”, 「手書き文字認識技術の過去・現在・未来」シンポジウム 講演論文集, pp.54-63, (1993).
- [8] 梅田 三千雄, “手書き漢字認識研究の歩み”, 電子情報通信学会技術研究報告, PRU, パターン認識・理解 95.278, pp.103-110, (1995).
- [9] UMEDA, Michio, “Advances in recognition methods for handwritten Kanji characters”, IEICE transactions on information and systems 79.5, pp.401-410, (1996).
- [10] 中野 康明, “文字認識・文書理解の最新動向[VI・完]—文字切り出しと文書理解—”, 電子情報通信学会誌, Vol.83, No.7, pp.576-580, (2000).
- [11] Alvy Ray Smith, “Color gamut transform pairs”, ACM Siggraph Computer Graphics, Vol.12, No.3, pp.12-19, (1978).
- [12] 西川 雅清, 宮本 拓弥, 松尾 賢一, “採点ミス発見支援システムの実現に向けた部分点及び採点記号の抽出と認識”, 電子情報通信学会関西支部学生会 第 19 回学生会研究発表講演会 講演論文集, D2-2,(2014).
- [13] 三宅 康二, “実用性の高い OCR の実現を求める基礎的研究”, 電子情報通信学会技術研究報告. PRMU パターン認識・メディア理解 106.605, pp.55-60 ,(2007).
- [14] 鶴岡 信治ら, “加重方向指数ヒストグラム法による手書き漢字・ひらがな認識”, 電子情報通信学会論文誌 D 70.7, pp.1390-1397, (1987).
- [15] Yann LeCun, Corinna Cortes, Christopher J.C. Burges, “THE MNIST DATABASE of handwritten digits”, <<http://yann.lecun.com/exdb/mnist/>>.
- [16] 中野 康明, “文字認識・文書理解の最新動向[IV]—文字パターンデータベース—”, 電子情報通信学会誌, Vol.83, No.5, pp.383-387, (2000).
- [17] 宮本 拓弥, 松尾 賢一, “答案採点支援システム実現に向けた重畳採点記号認識”, D-12-3, 電子情報通信学会総合大会講演論文集, (2013)