

採点ミス発見支援システムの開発Ⅱ ～汎用性を向上させる処理部の導入～

Development and Evaluation of Support System for Finding Scoring Error

三上徹朗[†] 松尾 賢一[†] 西川 雅清[†]
Tetsuro Mikami Ken'ichi Matsuo Masakiyo Nishikawa

1. はじめに

人手による記述式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、記述式答案の採点を支援するシステムの開発が検討されてきた。例として、テキストマイニングや機械学習を応用して短答式記述問題を自動採点する試み[5][6]がなされてきた。しかし、採点作業の支援をする有効なシステムの提案には至っておらず、コンピュータによる解答文字の意味・理解といった推論機能の技術的な困難さが浮き彫りとなった。そこで、採点済みの記述式答案に着目すると、自然言語の他に単体で意味を持つ部分点や採点記号が追記される。これに対して、認識機能としてオフライン手書き文字認識処理[7]を利用可能といえる。この要素技術を用いて、西川らは採点済みの答案上の採点記号と部分点の認識結果から点数の不一致を判定し、その箇所を採点ミスとして採点者に提示する「採点ミス発見支援システム」を開発した。

本システムでは、オフライン手書き文字認識処理により印刷された解答欄や鉛筆で書かれた解答文字が存在する答案に筆記された単一の手書き採点記号や部分点の認識が可能であった。そして、2名の採点者による模擬答案を用いた性能評価により採点ミス発見支援システムの実現性を示した[8]。しかし、多人数の被検者による性能評価が未確認であった。

また、性能評価で用いた答案は印刷されている独立した解答欄であったため、解答文字に対する採点記号は単一であったが、連続した解答欄が印刷された答案では、単一の採点記号同士が接触・重畳した採点記号となることが多い。これに対して、本システムの文字抽出部では、この接触・重畳した採点記号に対する分離・抽出は未対応であった。

本論文では、採点ミス発見支援システムの実用性の確認と汎用性を向上させる処理部の導入について述べる。まず、システムの実用性を明らかにするため被験者を増加させたシステムの性能評価実験を実施する。次に、採点作業を撮影した映像から採点記号の筆跡情報を取得する提案手法をシステムの処理部として導入し、前述の接触・重畳した採点記号に対する分離・抽出の実現性を実験結果より明らかにする。

2. 答案・採点ミスの定義

図 1 に示す模擬答案用紙 (高専入試解答用紙) には、答案用紙に事前に印刷されている「印刷領域」として、「解答欄」、「部分点記入欄」、「印刷文字」が存在する。これに対し、「解答文字」および「採点記号」と「部分点」を記入される。

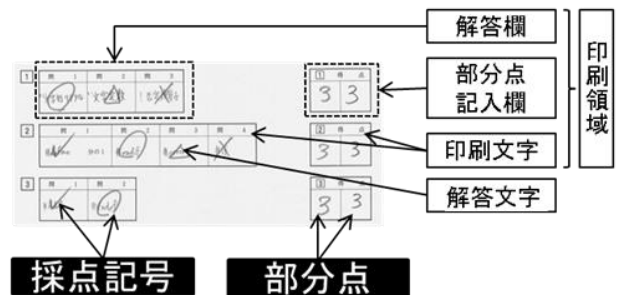


図 1 模擬答案用紙

試験では、この答案用紙の解答欄内に受験者は手書きで解答文字を記入する。採点作業では、まず解答文字に対し採点者が正誤の判定を行い、その正誤結果に対して採点記号を用いて解答文字付近に重ね書きする。最後に、各採点記号の配点に基づき部分点記入欄内に部分点を記入する。高専入試の採点作業は、採点記号の見誤りや見落としによる採点ミスの発生を防止するため、採点者に対して採点記号や部分点を枠内に正確に記入する制約を採点マニュアルで定めている。

ここで、採点作業の時系列で発生する採点ミス事例を表 1 内の I から V に示す。

表 1 採点作業上で発生する採点ミス事例

採点作業	事例
↓	I. 正答と誤答の判断を間違う「正誤判定誤り」
	II. 採点漏れなどの「記入誤り」
	III. <u>採点記号と部分点の「対応付け誤り」</u>
	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

ここで、下線を引いた表 1 の採点ミス III である「対応付け誤り」の答案例を図 2 に示す。図 2 の答案に対して、採点記号 '○' に 1 点、'/' に 0 点が配点されている。これより、図 2 の答案では、'○' が 3 つ、'/' が 1 つであることから部分点は本来 3 点となる。ところが、対応付け誤りでは部分点記入欄に 2 点が誤記入されていることがわかる。つまり、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対応付け誤りとなる。

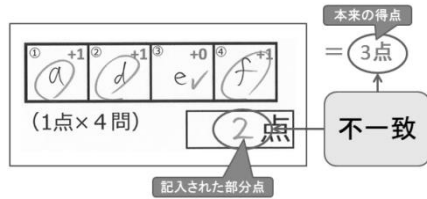


図 2 対応付け誤り例

3. 採点ミス発見支援システム

筆記者に制約を与えて筆記させた文字は「制約手書き文字」、筆記者に自由に筆記された文字は「自由手書き文字」と呼ばれる[9]. この自由手書き文字では、制約手書き文字と比較して、文字位置の検出や文字切り出しが非常に困難である。また、両者の文字を認識処理する場合、文字検出や切り出しが容易である制約手書き文字の方が文字認識技術の適用が容易である。よって、表 1 に示す全ての採点ミスを発見支援するシステムの実現には、手書きで筆記された解答文字が制約手書き文字であることが望ましいといえる。

これに基づき、採点者に印刷領域内に採点記号および部分点を筆記する制約を課し、両者の文字に対して制約手書き文字認識技術を導入することで、表 1 内の採点ミスⅢの「対応付け誤り」を発見する採点ミス発見支援システムの開発に至った。また、採点ミス発見支援システムに対して、2 名の被験者による評価実験の結果からその実現性を明らかにした。この採点ミス発見支援システム（以後、従来システム）の概要を図 3 に示す。従来システムは、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制約手書き文字認識結果間で不一致が生じたとき、答案内の「対応付け誤り」箇所を採点者に通知する機能を有する。

実際の作業では、従来システムの出力となる「対応付け誤り」が発生した箇所の通知に従って、採点者は誤り発生箇所を再チェックすることで対応付け誤りの有無を確認することになる。

従来システムは、図 3 に示す「文字抽出部」、「パターン認識部」、「ミス判定部」の 3 つの処理部から構成される。

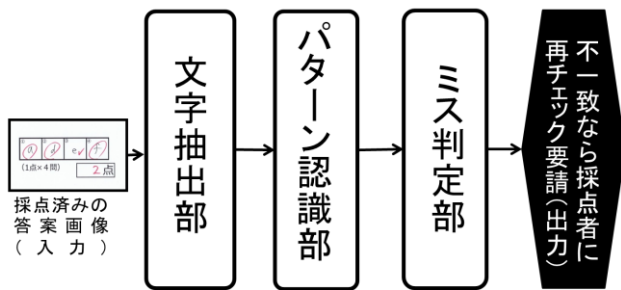


図 3 採点ミス発見支援システムの概要

赤色の筆記具で採点済みの答案を画像として入力し、「文字抽出部」で色情報を用いて答案画像から採点記号と部分点パターンを分離抽出する。「パターン認識部」において分離抽出された部分点パターンに対しては、「部分点認識」で手書き数字認識処理によって、0 から 9 の数字と

してカテゴリ分けすることで、答案上の部分点を得る。同様に、採点記号パターンに対しては、「採点記号認識」で手書き文字認識処理によって、4 つの採点記号にカテゴリ分けされる。カテゴリ分けされた採点記号の配点に基づき、最終的な合計点が算出される。

最後に、「ミス判定」で、パターン認識部で得られた部分点と合計点の正誤を判定する。誤判定であれば、不一致が生じた採点ミスの箇所を採点者に通知することで、採点者に再チェックを促す。以後において、従来システム内の各部の具体的な処理内容について詳細に述べる。

3.1 文字抽出部

図 3 内の文字抽出部内の処理を図 4 に示す。文字抽出部では、印刷文字、解答欄の枠線、枠線内に記入された解答文字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、採点記号および部分点パターンを抽出する。

採点作業が終了した答案用紙をスキャナによって RGB 形式で画像化する。図 4 に示すように、答案画像を文字抽出部の入力とする。

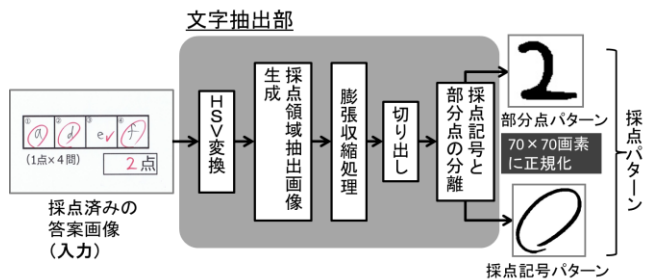


図 4 文字抽出部

答案画像の RGB 値を基に、色相 H、彩度 S、明度 V の 3 つの属性をもつ HSV 値に変換する[10]. 採点作業では、一般的に赤色の筆記具が用いられることが多いため、答案画像内の筆記具の色情報と同様の HSV 値域内の画素を「採点領域」、それ以外の紙面、印刷領域、解答文字である「背景領域」と定義する。予備調査による採点領域と背景領域の HSV 値域を表 2 に示す。[11]

表 2 採点および背景領域の 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

文字抽出部で採点領域と背景領域の分離に、答案画像内の画素の彩度 S が表 2 の採点領域と背景領域の各々の彩度 S の値域内であるか否かで分類する。さらに、分類した採点領域の画素に対して、赤色の筆記具の色相 H の値域内であるか否かでさらに再分類する。

最終的に、表 2 に示す採点領域の色相 H と彩度 S の値域内の画素を画素値'1'、それ以外は画素値'0'をもつ 2 値画像を生成する。この 2 値画像を「採点領域候補画像」と定義する。採点領域抽出画像内の採点領域のストロークが複雑

背景や採点領域と重畳する箇所では、混色の影響によって HSV 値の変動する。よって、文字抽出処理において重畳したストローク箇所において、ストロークの一部に欠損や断線が発生する。このストロークの欠損や断線については、採点領域候補画像に対する膨張収縮処理によって、ストロークを修繕する。

次に、採点領域候補画像の画素値‘1’をもつ各閉領域の面積を T とする。ここで、 $T < 50$ の閉領域は「雑音領域」とみなし、採点領域候補画像から除去する。なお、 $T < 50$ は、予備調査から決定した。雑音領域除去後の採点領域候補画像内の各閉領域を「採点パターン」と定義し、この採点パターンは縦横 70×70 画素の大きさに正規化され、文字抽出部の出力結果となる。すでに、図 4 に示した文字抽出部に対して、答案画像を入力に用いた予備実験は実施済みであり、採点パターン抽出率として 100%の結果が得られている.[8]

3.2 パターン 認識部

図 3 内のパターン認識部の処理を図 5 に示す。このパターン認識部には、文字抽出で得られた部分点および採点記号の 2 種類の採点パターンが入力される。

採点パターンは、パターンの重心による位置合わせがなされ、特徴抽出は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[12][13] (64 次元)、識別器には、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) を用いて得られた 2 種類の採点パターンの認識結果が出力される。

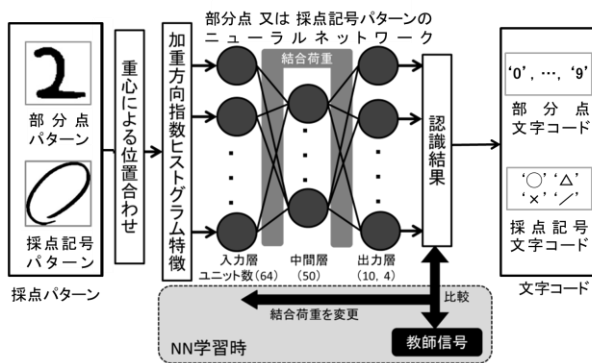


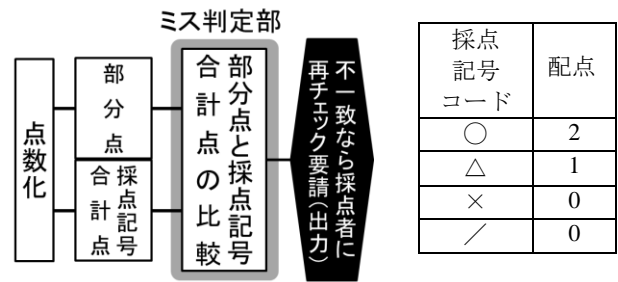
図 5 パターン認識部

この NN は、部分点と採点記号パターンで個別に構築されており、出力層の各ユニットに認識対象である採点パターンが文字コードに対応づけられる。

まず、部分点認識用 NN では、採点パターンの認識結果を 10 個の出力層ユニットに文字コードとして ‘0’ から ‘9’ を割り付け、採点記号認識用 NN では、採点記号の認識結果を 4 個の出力層ユニットに文字コードとして ‘○’, ‘△’, ‘×’, ‘/’ の割り当てる。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。

3.3 ミス判定部

図 3 内のミス判定部の処理を図 6 に示す。このミス判定部は、大問毎にパターン認識部で得られた 10 文字の部分点文字コード(‘0’, …, ‘9’)を点数化した部分点、ならびに 4 文字の採点記号文字コード(‘○’, ‘△’, ‘×’, ‘/’)を図 6(b)の配点表に従い点数化し、大問毎に合計した採点記号合計点を入力とする。ここで、ミス判定部では、大問毎における部分点と採点記号合計点を比較する。この比較結果より、部分点と採点記号合計点が不一致であれば、答案画像上で採点ミスの可能性あるとして大問の位置を採点者に提示し、再チェックを要請する。



(a) ミス判定部

(b) 配点表

図 6 ミス判定部

4. システム性能評価実験

従来システムは、これまで部分点と採点記号文字パターンの認識結果から、部分点と採点内容から得られた両得点間でミス判定を行い、採点ミスを採点者に提示することが可能であるかを、実験協力者 (採点者) 2 名の採点答案と個人辞書、ならびに、5 名の実験協力者で作成した汎用辞書を用いて、従来システムの実現性を明らかにした[8]。

これに対して、本論文では、採点ミス発見支援システムの汎用性を明らかにするために、新規システムとして左利きの筆記者を追加した実験協力者 6 名による汎用辞書を作成する。また、実験協力者(採点者)を 2 名から 10 名に増やし、模擬答案及び各々で個人辞書を作成する。

性能評価実験では、新規システムに対して従来システムと同様の評価実験を実施し、従来システムと新規システムとの性能の違いを実験結果で示すと同時に、実験結果に基づく新規システムの有用性を確認と実験結果による考察をする。

4.1 実験環境

性能評価実験で用いる答案用紙を図 7 に示す。図 7 の答案用紙は、実験協力者 (採点者) によって模擬的な採点作業で作成される。答案用紙には、大問が 10 問、各大問に 4 から 6 箇所の解答欄があり、合計 50 箇所の解答欄が存在する。また、答案用紙の右端に、大問ごとに部分点記入欄、解答用紙上部に総得点記入欄が存在する。この答案用紙への模擬採点では、赤色の筆記具を用いて全て筆記される。

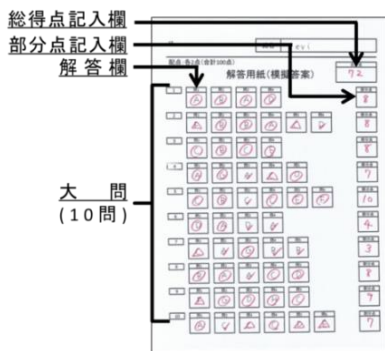


図 7 答案用紙例

採点作業は、解答欄に解答文字が記述された答案用紙と事前に印刷された模範解答用紙を採点者に配布し、採点者は模範解答用紙を参照しながら、解答文字に対する正誤を採点記号によって解答欄付近に筆記する。採点記号による採点作業終了後、大問ごとに採点記号に割り当てた配点の小計を部分点記入枠に筆記する。次に、部分点記入欄の全てに部分点の筆記が終了すれば、部分点を合計した点数を総得点記入欄に筆記する。

以上の採点作業は、10名の採点者 A~J で実施する。採点作業に用いられる模範答案は、全 1000 枚で 1 名あたり 100 枚である。手書きで筆記する採点記号は、‘○’、‘△’、‘×’、‘／’ の計 4 種類、部分点は ‘0’ から ‘9’ の 1 桁の数字のみとする。筆記場所について、解答欄や部分点記入欄の枠線との接触、重畳を許容しつつ、記号や数字の中心位置が記入欄外にならないよう採点者に事前に指示する。答案用紙は、イメージスキャナ (brother 製 justio MFC-6490CN) を用いて、解像度 300dpi で画像化する。

4.2 評価方法

認識部に用いた不特定多数の採点者の採点パターンを認識を対象とした汎用辞書と既知の採点者を対象とした個人辞書の 2 種類に分けてシステムの採点ミス提示精度の評価する。4.1 で説明した模擬答案を用いる。大問の数は合計 1 万個 (模擬答案 1 枚あたり大問 10 個 × 100 枚 × 10 人) であり、採点ミスは答案内に合計 162 個存在する。答案用紙の書式は既知とする。また、システムによる採点ミス提示は大問ごとに行う。

ここで、システムの採点ミス提示精度の指標として、再現率 R を用いる。再現率 R は、模擬答案に含まれる採点ミスである正解群 β [個] とシステムの提示により発見できた採点ミスの正提示 δ [個] から

$$\text{再現率 } R[\%] = \frac{\text{正提示 } \delta [\text{個}]}{\text{正解群 } \beta [\text{個}]} \times 100 \quad (1)$$

より算出する。この再現率 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提示を表す網羅性の指標である。

4.3 認識辞書

部分点である手書き数字パターンと採点記号である手書き文字パターンに対して、認識部の処理に必要な認識辞書

を個別に作成する。さらに、各辞書には、答案採点者の両者のパターンを辞書にした個人辞書、答案採点者以外の不特定多数の採点者のパターンを辞書にした汎用辞書を作成する。

汎用辞書は、模擬答案採点者以外の 6 名の筆者 (右利き 5 名、左利き 1 名) によって筆記された部分点パターン 1600 個 (‘0’ から ‘9’ 各 160 個)、および、採点記号パターン 640 個 (‘○’、‘△’、‘×’、‘／’ 各 160 個) を NN に学習させた認識辞書である。

これに対して、実際の採点作業では、一人の採点者のみで採点が行われることもある。このような状況では、採点者だけの部分点パターンと採点記号パターンによる認識辞書である個人辞書を作成することで、汎用辞書と比較して認識精度が向上することが予想される。

そこで、採点者だけで筆記された部分点パターン、および、採点記号パターンを NN に学習させた認識辞書を個人辞書として作成する。個人辞書の作成には、部分点パターン 1600 個 (‘0’ から ‘9’ 各 160 個)、採点記号パターン 640 個 (‘○’、‘△’、‘×’、‘／’ 各 160 個) を用いる。

4.4 汎用辞書を用いたときの再現率

システム動作時の採点ミス提示結果の画面例を図 8 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、システムの採点ミス提示精度を評価する。

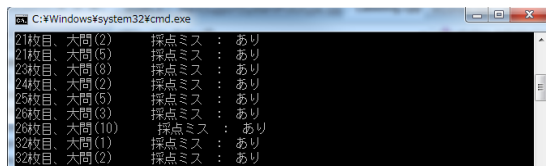


図 8 採点ミス提示結果の画面例

表 4 に 10 名 (A~J) の採点者により採点された 1000 枚の模擬答案に対して、汎用辞書を用いたときの再現率 R を示す。また、各採点者の模擬答案で発生した正解群 β (採点ミス) とシステムによる出力 (採点ミスの可能性がある箇所の提示数)、それに対する正提示 δ を示す。

表 4 汎用辞書での再現率

採点者	正解群 β [個]	出力 [個]	正提示 δ [個]	再現率 R [%]
A	22	110	22	100.0
B	3	160	3	100.0
C	22	195	20	90.9
D	9	300	8	88.9
E	51	161	48	94.1
F	10	209	10	100.0
G	14	119	13	92.9
H	4	300	3	75.0
I	10	172	9	90.0
J	17	108	17	100.0
合計	162	1834	153	94.4

表 4 の結果より、1000 枚の模擬答案に対して、採点者 A から J によって採点した結果、採点者各々の答案において、正解群 β である採点ミスが存在し、その合計は 162 個であることがわかる。この採点済みの答案データに対して、汎用辞書を用いた新規システムは、1834 個の解答欄で採点ミス箇所を提示している。この提示した箇所内、153 個において正確な採点ミスの箇所である正提示 δ より、式(1)の再現率 R は、全答案データで平均 94.4%であった。

採点者別での再現率 R は、4 名で 100%、8 名が 90%以上であった。

これらの結果から、不特定多数で作成した汎用辞書を用いても、90%以上の再現率が得られることが明らかとなった。しかしながら、左利きの採点者 H で再現率 R が 75%と最低となった。これは、汎用辞書が右利きの筆記者に偏って作成されたことが原因であるといえる。

4.5 個人辞書を用いたときの再現率

4.4 同様のデータに対し、個人辞書を用いたときの再現率 R を表 5 に示す。

表 5 個人辞書での再現率

採点者	正解群 β [個]	出力 [個]	正提示 δ [個]	再現率 R [%]
A	22	127	22	100.0
B	3	63	3	100.0
C	22	149	22	100.0
D	9	190	7	77.8
E	51	179	49	96.1
F	10	202	10	100.0
G	14	128	13	92.9
H	4	101	4	100.0
I	10	205	8	80.0
J	17	219	17	100.0
合計	162	1563	155	95.7

表 5 より、個人辞書を用いた新システムでは、正解群 162 個に対して、155 個の採点ミスの箇所が提示した。その結果、再現率 R は、全答案データで平均 95.7%となった。汎用辞書での再現率の平均 94.4%を上回る結果であった。また、採点者別の再現率 R は、6 名において 100%であり、その他の採点者も 80%以上の結果となった。汎用辞書で再現率 R が 75%と最低となった採点者 H の再現率 R が、個人辞書では、100%であった。よって、採点者が既知であるとき個人辞書を用いることで再現率 R が向上するため有用であるといえる。

4.6 汎用性に関する考察

従来のシステム[8]は、被験者 2 名の答案データに対して、2 名の採点者で作成した個人辞書、および、それ以外の採点者 5 名で作成した汎用辞書を用いて、両者の再現率を明らかにし、その再現率の結果からシステムを用いた採点ミス発見支援の可能性を明らかにした。しかしながら、データ数の少なさから汎用性まで明らかにするに至らなかった。

そこで、新規システムでは、筆記者の増員によるパターン認識部の精度向上と被験者の増員による汎用性の向上を

目的とした。汎用辞書、個人辞書の再現率 R を含め、従来および新規システムの再現率 R の結果を表 6 に示す。

表 6 従来と新規システムの再現率 R [%]

	再現率 R [%]	
	汎用辞書	個人辞書
従来システム	97.5	97.5
新規システム	94.4	95.7

表 6 に示すように新規システムでは、汎用辞書で再現率 R は平均 94.4%、個人辞書で平均 95.7%の結果が得られた。

新規システムでは、被験者を増員したことで、個人辞書・汎用辞書ともに従来システムと比べて本論での再現率 R は低下した。また、再現率 R は、汎用辞書よりも個人辞書の方が高く、採点者別でも再現率 R が 100%の人数が多い結果が得られた。

つまり、汎用辞書よりも個人辞書を使用する方が、採点ミス支援の精度が高いことを表し、結果として採点者による再チェック箇所が減少することを意味する。ゆえに、採点者が既知であるとき、その採点者の個人辞書を使用することが本システムの汎用性を高めるといえる。

逆に、採点者が未知であるとき、汎用辞書を使用せざるを得ないため、再現率 R の低下や出力数増加によって再チェック箇所の負担増加が予想される。これを抑止するために、未知の採点者による採点開始当初に、汎用辞書を利用しながら、採点作業を進める過程で常時採点パターンを取得し個人辞書を作成しつつ、作業途中で汎用辞書から個人辞書に切り替える処理を導入することが考えられる。

5. 汎用性を向上させる処理の導入

4.6 で述べたようにシステムの汎用性を向上させるためには、パターン認識部の再現率 R を高めることが重要である。

しかしながら、従来および新規システムにおいても、採点記号や点数の記入位置に制約のある答案画像が対象であった。実際の一般的な記述式答案の採点では、採点記号や部分・合計点数の記入位置の制約を課さず、自由記述であることが多い。よって、採点作業において、図 9 に示す採点記号同士の接触や重畳が多く発生することになる。このような文字を「接触重畳採点記号」と定義する。

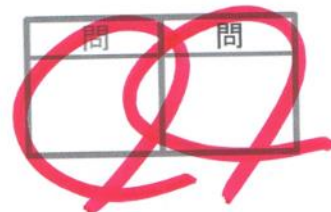


図 9 接触重畳採点記号

この接触重畳採点記号に対して、現状の採点ミス発見支援システム内の文字抽出部では、接触重畳採点記号の抽出は可能であるが、接触重畳採点記号を 1 文字の採点記号単位の分離抽出には対応できない。

仮に、接触重畳採点記号のまま抽出されて、次の採点記号認識部に引き渡されると、採点記号の誤認識や合計点の算出に誤りが生じ、結果的に再現率 R が低下する。

この問題は、本システムの汎用性を低下させる要因であることから、採点記号同士の接触や重畳に対して新たな処理を導入することは、汎用性向上に向けて必須といえる。

ここで、接触重畳採点記号のリジェクト処理の導入も考えられるが、これは再現率 R を低下させてしまうことにつながる。よって、再現率 R と汎用性を向上させるため、接触重畳採点記号を 1 文字単位の採点記号に分離抽出する処理の導入が必要となる。ここで、採点記号の書き始めから書き終わりまでの筆点座標を筆跡情報として利用可能であれば、一文字単位での確実な記号切り出しや記号の筆跡情報の取得が可能となる。そこで、汎用性を向上させる処理として、本システムの抽出部において、筆跡情報の取得を可能にする処理を追加したシステムの改良が考えられる。

この筆跡情報の取得が可能になれば、これまでの処理に新たに接触重畳採点記号の分離抽出にも対応できるように汎用性を高めることが可能になる。また、これまでの本システムの抽出した採点記号パターンと辞書とのパターンマッチングによるオフライン文字認識から、筆跡情報を用いたオンライン文字認識の処理を追加できるため高い認識率が期待できる。よって、再現率も飛躍的に向上が可能になる。そこで抽出部への追加処理として、カメラで撮影した採点作業から採点記号の筆跡情報とそれに基づく採点記号パターンを生成する手法を提案する。

提案手法では、カメラで採点作業を撮影した採点映像の各フレームから、色情報を用いて採点記号の部分ストロークを抽出する。次に、筆記開始から終了までの部分ストロークの重心座標を筆跡情報として算出する。この筆跡情報を基に採点記号パターンを生成する。この採点記号パターンをオンラインパターンと定義し、すべての採点記号の筆記が終了した答案画像から従来手法で抽出された採点記号パターンをオフラインパターンと定義する。

最終的に、提案した手法による採点記号パターンの筆跡情報の取得精度については、筆跡情報を基に生成した採点記号パターンに対して、4.3 で述べた認識辞書を用いた認識実験から得られた認識率によって明らかにする。更に得られた認識率から今後の接触重畳採点記号文字の分離抽出の実現性について考察する。提案手法の詳細と実験内容について次章で述べる。

6. 提案手法

提案手法の処理の概要を図 10 に示す。まず、カメラ (CIC 製: EVI-HD1) を用いて、採点作業を撮影する。カメラの設置条件を図 11 に示す。カメラは、筆記者から向かって左隣、答案から 100mm 離し、接地面から 70~100mm の高さの台に設置し、カメラを接地面に対して 30° に傾ける。

設置したカメラを用いて、解像度:1920×1080pixel でフレームレート:30fps で採点作業を撮影する。図 10 に示した追加処理の詳細を以降で述べる。

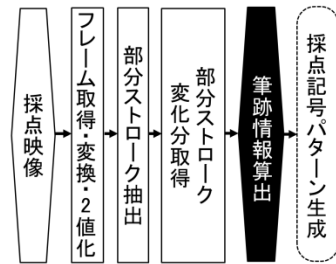


図 10 提案手法の処理の概要

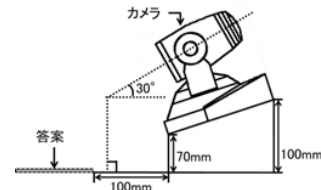
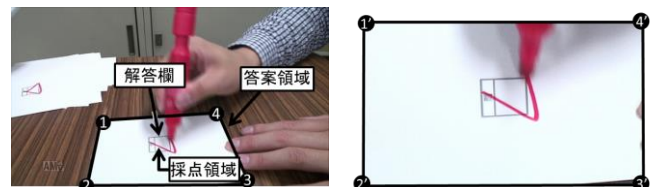


図 11 カメラの設置条件

6.1 答案画像と採点領域抽出画像

フレームレート:30fps で N 秒間採点作業を撮影した採点映像から $30 \times N$ 枚のフレームを取得する。各フレームは、図 12(a) に示す解答欄に対して単一の採点記号を筆記する採点作業が撮影されている。採点作業フレームでは、答案領域が奥行方向に歪んだ台形状で撮影されている。このフレーム内の台形状である答案領域の四隅の座標値を用いて、真上から見た答案の状態に切り出された図 12(b) の答案領域を取得する。



(a) 採点作業フレーム (b) 答案領域
図 12 採点作業フレームからの答案画像の取得

次に、答案画像から表 6 に示す HS 値域をもつ画素を採点領域として画素値 '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作成する。

表 6 採点領域の HS 値域

	色相 H	彩度 S
採点領域	0~10, 161~180	70~255

6.2 部分ストロークと部分ストローク変化領域

図 13 の採点領域抽出画像は、採点記号の部分ストローク、および、筆記者の手や筆記具などの雑音領域が採点領域として画素値 '1' (図 13 内では黒画素) が与えられている。この採点領域抽出画像から、雑音領域を除去することで採点記号の部分ストロークのみを抽出する。

まず、採点領域画像内の採点領域に対して、収縮膨張処理によって幅の狭い採点領域を除去したマスクプレーンを

作成する。つまり、図 13 においては、幅の狭い部分ストロークの領域が除去され、雑音領域のみが存在するマスクプレーンを得る。次に、採点領域抽出画像内のマスクプレーンが一致する領域を除去することで、採点記号の部分ストローク領域のみが存在する $30 \times N$ 枚の採点領域抽出画像が得られる。

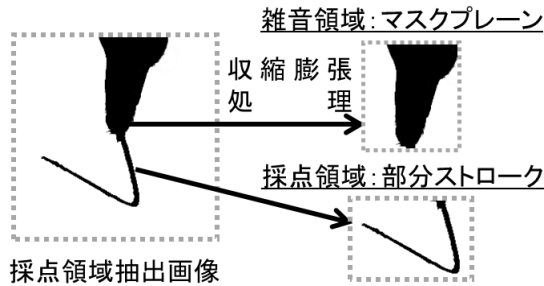


図 13 採点領域抽出画像

得られた採点領域抽出画像の前後で差分を取得し、連続する部分ストローク領域の変化分を示す変化領域を取得する。

6.3 筆跡情報算出、採点記号パターン生成

図 14 に示す変化領域付近において、X 軸、Y 軸方向に変化領域における画素数(図 14 内では黒画素)の総和を射影したヒストグラムの重心位置 x 、 y を求め、変化領域の重心座標 $G(x,y)$ を算出する。

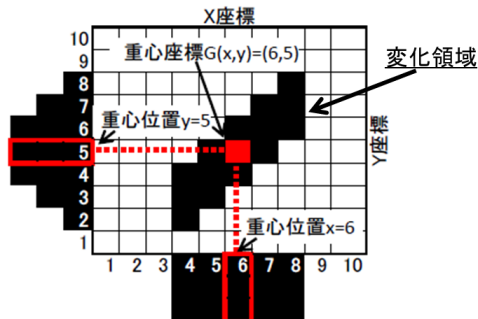


図 14 変化領域の重心座標 $G(x, y)$ の算出

この重心座標 $G(x,y)$ を全ての变化領域において算出する。ここで、一つの採点記号ストロークの筆記開始フレーム時間 s から筆記終了フレーム時間 e であるとき、このフレーム時間内の重心座標系列 $HI(G(x,y),t) (t=s, \dots, e)$ を筆跡情報と定義する。このとき、連続する筆跡情報の 2 点間である $HI(G(x,y),t)$ と $HI(G(x,y),t+1) (t=s, \dots, e-1)$ を全て直線で結線したパターンを生成する。このパターンをオンラインパターンと定義し、'○'、'△'、'×'、'/' のオンラインパターン例を図 15 に示す。なお、現状のオンライン処理部では、ペンのアップ・ダウンの検出機能が未実装であるため、図 15(c) に示す '×' のオンラインパターンは 1 画目の書き終わりと 2 画目の書き始めの筆跡情報が結線された形状となる。

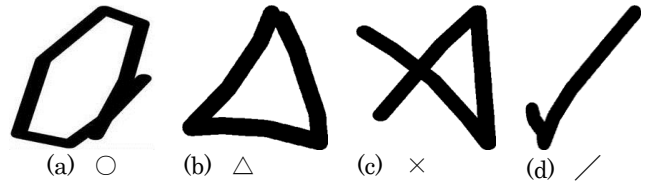


図 15 オンラインパターン例

また、すべての採点記号の筆記が終了した答案領域から従来手法で採点記号パターンを抽出する。これをオフラインパターンと定義する。

7. オンライン処理部に対する性能評価実験

従来のシステムでは、オフラインパターンの抽出に図 4 の文字抽出部を用い、スキャナ入力による答案画像を対象としていた。しかし、将来、システムに対して図 10 のオンライン処理部の導入した際、文字抽出部はカメラ入力による答案領域を対象としてオフラインパターンの抽出となる。そこで、従来手法がカメラ入力による歪みや採点領域の色味の変動が発生する答案領域を対象とした時、オフラインパターンの抽出処理の流用が可能か否か明らかにする。

また、オンライン処理部で取得した筆跡情報で生成したオンラインパターン形状を目視によって確認する。そして、オフライン文字認識の観点からオフラインパターンとオンラインパターンの認識精度を調査し、認識精度の差異を明らかにする。

実験環境を図 16 に示す。実験では、解答欄 1 つが中央に配置された用紙を用い、筆記者は赤ペンを用いて図 17 の解答欄付近に採点記号 '○'、'△'、'×'、'/' のどれか 1 つ筆記する風景を「採点映像」とする。「採点映像」として、14 名の右利き筆記者から各採点記号 700 回、合計 2800 回の筆記を撮影する。

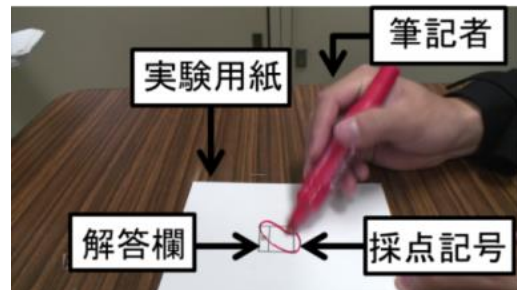


図 16 カメラから見た実験環境

7.1 オフラインパターン抽出精度評価実験

カメラ入力はスキャナ入力と比較して、答案画像に歪みの発生や照明状況によって色味の変動が起こる。よって、カメラ入力による答案領域から、図 4 の従来手法の文字抽出部によるオフラインパターンの抽出は困難となる可能性がある。そこで、採点映像内の答案領域から従来手法の抽出によりオフラインパターンの抽出精度を評価する。

オフラインパターン抽出結果を表 7 に示す。断線や欠損がないオフラインパターンを「抽出数[個]」として数え、各採点記号パターン数[個]に対する抽出数[個]の割合を「抽出率[%]」とする。

表 7 オフラインパターン抽出結果

採点記号 パターン	採点記号 パターン数[個]	抽出数 [個]	抽出率 [%]
○	700	700	100.0
△	700	698	99.7
×	700	700	100.0
/	700	699	99.9
合計	2800	2797	99.9

表 7 の実験結果から、合計の抽出率は 99.9% となった。この抽出結果から、歪みや変動する色味の発生するカメラ入力による答案領域においても従来手法の抽出率に変化がないことが確認できた。よって、図 4 の従来システムの文字抽出部は、カメラ入力による答案領域に流用可能であるといえる。

7.2 オンラインパターン形状評価実験

6.1 で撮影した「採点映像」から開発したオンライン処理部を用い、筆跡情報を取得する。そして、筆跡情報間を直線で結線し、オンラインパターンを生成する。オンラインパターンが筆記した採点記号パターン「○」、「△」、「×」、「/」の形状を再現できていると目視で確認できた場合を「良形」とする。なお、オンライン処理部にはペンのアップ・ダウンを検出する機能が未実装である。そのため、採点記号「×」は、1画目と2画目のストロークが連結された形状を良形とする。各採点記号の700回の筆記に対する良形数の割合を「良形率[%]」として表 8 に示す。

表 8 良形率

良形率[%](良形数[個])				
○	△	×	/	平均
81.4(570)	95.2(667)	93.4(654)	97.7(684)	91.9

表 8 より、「△」は 95.2%、「×」は 93.4%、「/」は 97.7% と高い良形率を示した。一方で「○」は 81.4% と他の採点記号と比べて良形率が低下した。

7.3 採点記号パターン認識精度評価実験

良形のオンラインパターンと抽出したオフラインパターンの認識精度評価実験を実施する。

7.1 より 7 名分を学習パターンとして 3.2 と同様に特徴抽出し、採点記号認識用 NN を用い認識用辞書を作成する。

また、学習パターンと別の 7 名分の採点記号パターンを未知パターンとして辞書で認識させる。辞書に対する学習パターンと未知パターンの認識率を表 9 に示す。

表 9 認識率

パターン種別	入力種別	認識率[%]				
		✓	○	×	△	平均[%]
オフラインパターン	学習	100	100	100	99	99.8
	未知	99.4	100	100	100	99.9
オンラインパターン	学習	100	100	99.7	100	99.9
	未知	78.2	97.3	77.2	100	88.2

オフラインパターンの辞書に対する学習パターンの認識率は、平均 99.8% であった。また、辞書に対する未知パターンの認識率は平均 99.9% であった。一方で、オンラインパターンの辞書に対する学習パターンの認識率は、平均 99.9% であった。また、辞書に対する未知パターンの認識率は平均 88.2% であった。両者の認識率から学習パターンにおいて認識率の差が 0.1% と極めて小さいが、未知パターンにおいては差が 11.7% と大きいことが分かる。

7.4 性能評価実験に関する考察

スキャナ入力による答案画像からのオフラインパターンの抽出率は 100% であった[8]。これに対して、7.1 のオフラインパターン抽出精度評価実験においてカメラ入力による答案領域からのオフラインパターンの抽出率は 99.9% であった。カメラ入力による答案領域には歪みやスキャナ入力の答案画像と異なる色味が生じたが、入力形式の違いによる抽出精度の差は 0.1% であった。よって、図 4 の従来システムの文字抽出部は、カメラ入力による答案領域に流用可能であるといえる。

また、接触重畳採点記号を 1 文字単位の採点記号に分離抽出する処理の導入のため、カメラによる映像から筆跡情報を取得するオンライン処理部を開発し、オンラインパターンを生成した。7.2 のオンラインパターン形状評価実験より、良形は 91.9% であった。

従来システムの文字抽出部とオンライン処理部を比較すると、良形は 8.1% 低下した。これは、部分ストロークの抽出過程において、筆記者の手によって部分ストロークが隠蔽によって抽出できず、正確な部分ストローク変化領域の取得ができなかったためである。よって、オフラインパターンを形成する筆跡情報数が不足し、元の採点記号パターンの形状を示せず良形に至らなかった。今後、部分ストロークが筆記者の手によって隠蔽された時、隠蔽部分の筆跡情報をオフラインパターンから補完するフィードバック処理の開発が必要である。

開発したオンライン処理部を用いることで採点記号パターンの書き始めや書き終わりといった筆跡情報が可能となる。よって、それらの筆跡情報から接触・重畳発生前後で従来の文字抽出部において取得される採点記号パターンの筆記順を決定でき、個々の採点記号パターンの切り分けに有効であると考えられる。

最後に 7.3 の採点記号パターン認識精度評価実験より、オフラインパターンでは、学習パターンにおいて平均 99.8%、未知パターンにおいて平均 99.9% という認識率が得られた。また、オンラインパターンでは、学習パターンにおいて平均 99.9%、未知パターンにおいて平均 88.2% という認識率が得られた。

結果より、取得した筆跡情報から再現されたオンラインパターンは、各採点記号の分類が可能な特徴を有しているといえる。しかし、オフラインパターンとオンラインパターンの認識率を比較すると、学習パターンにおいて認識率の差が 0.1% と極めて小さいが、未知パターンにおいては差が 11.7% と大きいことが分かる。筆跡情報を結線して生成したオンラインパターンの形状は、各採点者間で算出できる筆跡情報数によって大きく変動する。また、オンライン処理部ではペンのアップ・ダウンを考慮した筆跡情報の切り分け処理が未実装であるため、「×」のオンラインパ

ターンは 1 画目の書き終わりと 2 画目の書き始めの筆跡情報を結線している。よって、オンラインパターン ‘x’ は各採点者の書き順の違いにより形状の変動する。そのため、オンラインパターンにおいて学習パターンによって作成された認識用辞書の有する特徴と未知パターンの特徴に隔たりが存在したため、認識率が低下したと考えた。

今後は、筆跡情報の高精度化やペンのアップ・ダウンの判定処理を開発する必要がある。また、オンラインパターンに対するオフライン文字認識ではなく、筆跡情報に対するオンライン文字認識によりオンラインパターンの形状の変動の有無に関わらず、高い認識精度が得られると考えた。

8.おわりに

筆者らは、これまでに答案画像から採点記号パターンと部分点パターンを抽出し、両パターンに対する認識処理の結果によって、両者間の点数に不一致があれば採点ミスの可能性が高い箇所を採点者に提示する「採点ミス発見支援システム」を開発した。これを従来システムとし、模擬答案の作成者ならびに採点者を増加させた新規システムに対して、性能評価実験によって実用性を確認した。また、新規システムの汎用性をさらに向上させるために、抽出部において採点映像から採点者が筆記した採点記号の筆跡情報を取得する処理部の導入を実施した。

従来システムでは、汎用辞書ならびに個人辞書を用いたシステムの性能評価の尺度である再現率 R は共に 95.7%であった。これに対して、採点者数増加させた新規システムにおける再現率 R は、汎用辞書で 94.4%、個人辞書で 95.7%であった。この結果から、被験者の増加によって再現率 R の低下が見られたが実用上問題のないレベルの低下率であった。また、採点者が既知であるとき、汎用辞書よりも個人辞書を用いた方が採点ミスの発見支援の精度が高いことがわかった。

次に、実際の採点で見られる採点記号同士の接触や重畳したパターンに対して新規システムが対応できるように、汎用性をさらに高める処理の追加として、採点作業をカメラで撮影した映像から筆跡情報を取得可能にする処理を提案手法として導入した。提案手法の有効性を後述する 3 つの実験で検証した。

一つ目は、スキャナ入力とカメラ入力での採点文字パターンにおける色情報の差異を、スキャナ入力およびカメラ入力の両者の答案画像から採点記号パターンの抽出実験により明らかにした。二つ目は、カメラ入力によって得られた採点映像中の単一採点記号の筆記作業から採点記号の筆跡情報の取得精度をオンラインパターン形状評価実験により明らかにした。三つ目は、得られた採点記号の筆跡情報から再度採点記号パターンを生成し、得られた筆跡情報で正確な採点文字パターンが再現されているかを新規システム内の認識部を用いた認識実験によって明らかにした。3 つの実験結果を通じて、カメラ入力による採点文字パターンの抽出ならびに筆跡情報の取得が可能であるといえる。

よって、カメラで撮影した採点作業から得られた筆跡情報を利用して接触・重畳した採点記号の分離・抽出の実現性を 3 つの実験結果から示したといえる。

今後、カメラ入力によって採点作業を撮影しながら、採点文字パターンの抽出と筆跡情報を取得する処理を新規シ

ステムの抽出部に導入し、単一採点記号パターンでなく接触重畳採点記号パターンを含む答案画像に対する採点ミス発見支援システムへと改良を加えていくことが今後の課題である。

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] 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).
- [5] 中島 功滋, “短答式記述答案の採点支援ツールの開発と評価”, 言語処理学会 第 17 回年次大会 発表論文集. コンピュータと教育研究会報告, 95,111,pp.49-54,(1995).
- [6] 石岡 恒憲, 亀田 雅之, 劉 東岳, “人工知能を利用した短答式記述採点支援システムの開発”, 電子情報通信学会言語理解とコミュニケーション 第 3 回自然言語処理シンポジウム. 電子情報通信学会技術研究報告, pp.87-92,(2016).
- [7] 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).
- [8] 西川 雅清, 松尾 賢一, “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～”, FIT2015 (第 14 回情報科学技術フォーラム) ,(2015)
- [9] 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).
- [10] Alvy Ray Smith, “Color gamut transform pairs”, ACM Siggraph Computer Graphics, Vol.12, No.3, pp.12-19, (1978).
- [11] 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).
- [12] 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).
- [13] 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).