

# 手書きノートのページセグメンテーションによる 学習活動の分析

李 柏毅<sup>1,a)</sup> 峰松 翼<sup>2,b)</sup> 谷口 雄太<sup>2,c)</sup> 大久保 文哉<sup>2,d)</sup> 島田 敬士<sup>2,e)</sup>

**概要:** 手書きノートは学習の重要な記録であり、授業で学んだ内容の記録、知識の整理、学習の振り返りやさらなる深化などの用途で利用される。また、教員は学生のノートを確認して、学習状況を把握することができる。手書きノートを、テキスト、数式、図、表など、さまざまな領域に分割して、定量化を行うことで、ノート作成状況を評価したり、学生間での比較を行ったりすることが可能になる。自動評価方法があれば、教師の読むことの負担が減り、学生にとっても参考になる。本研究では、手書きノートのページセグメンテーション結果と学習活動の関係分析についての報告と、ディープラーニングに基づく、ページセグメンテーションの自動化に関する検討状況を報告する。

## 1. はじめに

学習者が自ら学習内容について記述したノートは、授業内容の記録や知識の整理、学習内容の振り返りなどに利用されるため、学習において重要な要素である。ノートには、学習者自身が学習内容をテキストや数式、図などを用いてまとめた情報が記述されている。このような情報が記録されたノートは学習者自身が知識の整理や学習内容の振り返りに利用できる点で有用であるだけでなく、ノートの記述内容の分析による学習状況の評価においても有用である。しかし、学習者により作成されるノートは手書きした紙の文書に代表される記述内容が構造化されていないものがほとんどであり、すべての学習者のノートを手動で評価することは、教員にとって大きな負担である。さらに、紙の文書などは非電子データであるため、機械による自動解析に適していない。したがって、手書きノートなどをノートの写真を撮るなどして得られた電子データから情報を自動で抽出し分析する手法が必要になる。

手書きノートの分析として、ノートのページ内に含まれるテキスト、数式、図などのコンテンツの定量化が必要になる。テキストの文字数はノートの作業量を反映することができ、文字数が多いほど、より多くの内容を含んでいる可能性がある。また、数学や物理など、図を描く必要があ

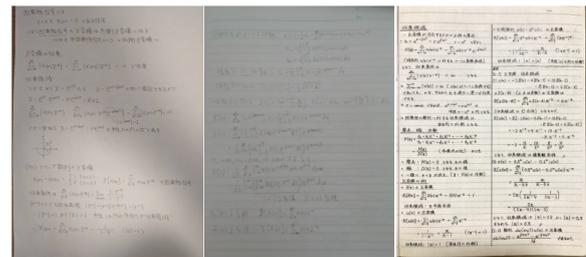


図 1 手書きノート

る科目の場合、図の量も同様にノート分析の指標になる。手書きノートに含まれるテキストを分析するために、光学式文字認識 (OCR) を利用することができるが、数式や図領域を同様に抽出する手法が必要になる。従来、手書きノートに含まれるコンテンツを分類することは困難であったが、近年の深層学習等の機械学習手法の発展により、画像内に含まれている各領域の属性を推定するセマンティックセグメンテーションの性能が向上してきている。

そこで、本研究では、深層学習による領域分割手法であるページセグメンテーションを手書きノートの定量化に適用し、ノートの記述量と学生の学習活動データとの相関分析を行う。本実験において、ページセグメンテーションの結果から文字分割を行い、文字数と図領域の面積に基づくノートの自動評価で得られた結果から、学生の教科書閲覧時間とテストの点数について相関分析を行う。相関分析により、ノートの記述量が学生の学習状況の関係を検証する。

## 2. 関連研究

「ノートをとる」という行為には、教科書や授業で学習

<sup>1</sup> 九州大学 大学院システム情報科学府 情報知能工学専攻 修士課程

<sup>2</sup> 九州大学 大学院システム情報科学府 情報科学研究所

a) li@limu.ait.kyushu-u.ac.jp

b) minematsu@ait.kyushu-u.ac.jp

c) yuta.taniguchi@ait.kyushu-u.ac.jp

d) fokubo@ait.kyushu-u.ac.jp

e) atsushi@ait.kyushu-u.ac.jp

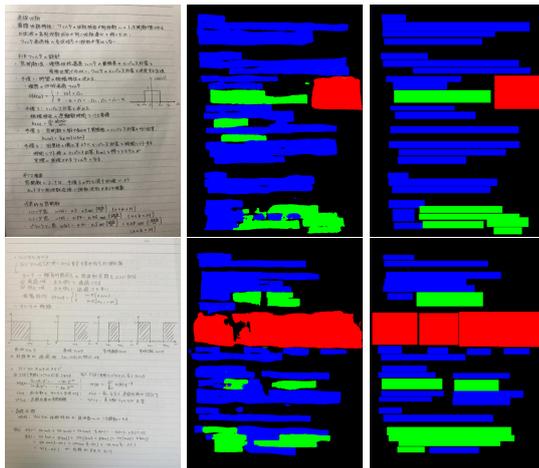


図 2 U-Net のセグメンテーション. 左から入力画像, 領域分割結果, 正解画像. 青: テキスト 赤: 図 緑: 数式.

したことを整理することが含まれ, ノートテイキング量とテストの点数の間に強い相関があることが報告される [1] など, ノートに関わる学習行動と学習効果についての研究が注目されている [2], [3].

ノートの記述内容の分析では, 紙に授業内容を記述させ, その内容を教員が分析し, 分析結果に基づいて, 成績やテストの点数などと相関分析を行う方法が提案されている [2]. 文献 [2] の学習分析支援システムでは, 紙のレポートをスキャンし, 学生の手書きの図を抽出することができる. 学生の最終的な GPA との相関関係は, 抽出された図のファイルサイズやレポートの提出の有無によって分析される. しかし, 特定のレポート用紙に依存した画像抽出手法となっている. また, [3] において, 学生間でのノート相互評価を行い, 内容のわかりやすさや見やすさを評価している.

一方で, 手書きノートの自動的な分析を見据えて, 手書きノートの文字領域や図領域を自動的に抽出する手法が提案されている [4]. 文献 [4] では, 画像処理を用いて, 文字領域および線や矩形などの図形領域を抽出する. 近年では, テキストや数式, 図などが含まれている文書に対応した手法として, 畳み込みニューラルネットワークを用いたページセグメンテーション手法が提案されている [5], [6]. 手書きノートに対しても, テキスト領域を検出するニューラルネットワークが提案されているが, 文字数などの詳細な分析方法は提供されていない. そこで本研究では, 画像処理と深層学習を組み合わせて, より詳細なページセグメンテーションを行い, 学生の学習活動データとの相関分析を行う.

### 3. 手書きノートのページセグメンテーション

本研究では, 手書きノートに記述されている図やテキスト領域等の種類ごとに分析するために, 手書きノートの画像の領域を自動的に分類するニューラルネットワークとし

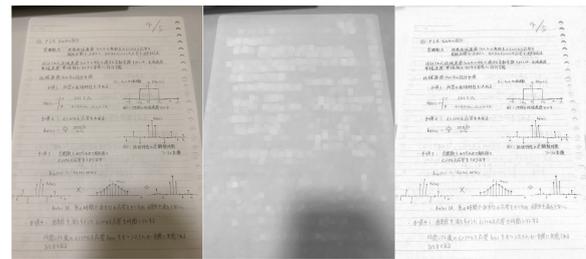


図 3 影の除去. 左から入力画像, 背景画像  $B'$ , 影除去の結果.

て U-Net を用いる [7]. また, テキスト領域に含まれている文字数を数えるために, テキスト領域の行ごとの抽出を行い, 抽出された各行ごとに, 文字分割を行う.

#### 3.1 U-Net を用いたページ内コンテンツの領域分割

本研究では, U-Net によって手書きノートの各ページ内に含まれる図や数式, その他の文章の領域に分割する. U-Net は, ニューラルネットワークの入力層に近い層で得られた特徴表現を Skip connection によって出力層に近い層の特徴と統合することで, 効果的に領域分割を行う手法として領域分割タスクにおいて利用されている.

図 2 は, 入力画像, U-Net によるページ内コンテンツの領域分割の結果, 対応する正解画像の例を示している. これらの結果から, U-Net によりページ内コンテンツの領域分割が可能であることが分かり, ページ内に含まれる図が占める面積を定量化できることを示唆している. 本研究では, テキストに関する文字数について分析を行うために, U-Net による領域分割の結果に加えて, モルフォロジー変換を利用したテキスト領域の行単位の抽出を行い, 文字数計測のための文字分割の前処理を行う.

#### 3.2 テキストの行単位の抽出

本研究で取り扱うノートは, 一般的な白背景のノートであるため, 記述領域は白以外の黒や赤色などの色を含む画素で構成される. そこで, 画像に含まれるエッジ特徴に着目し, テキスト領域に含まれる文字数を計測するための前処理として, テキストの行単位の抽出を行う. ただし, 手書きノートは学習者によって異なる照明環境で撮影されるため, 手書きノート内に影が映りこむことが多い. 提案手法では, 影による検出性能の低下を避けるために, ページ画像内の局所領域ごとに, 画素値の最大値・最小値を抽出する max, min フィルタを用いて影を除去する.

影がある場合, 画像の画素は大きく分けて, 影のない背景画素, 影のある画素, 手書き文字を構成する画素の 3 種類に分けられると仮定すると, それぞれの輝度値  $P_o, P_s, P_w$  の関係は次のように考えられる.

$$255 \geq P_o > P_s > P_w \geq 0 \quad (1)$$

まず, フィルタサイズ  $20 \times 20$  の max フィルタを使っ

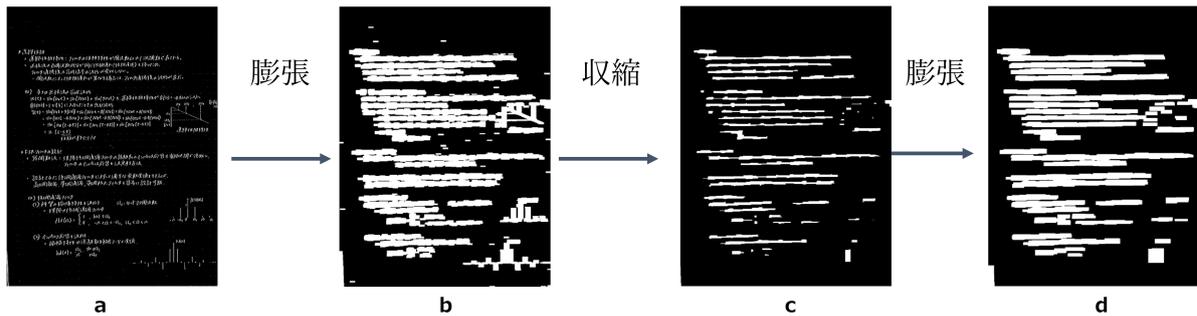


図4 モルフォロジー変換

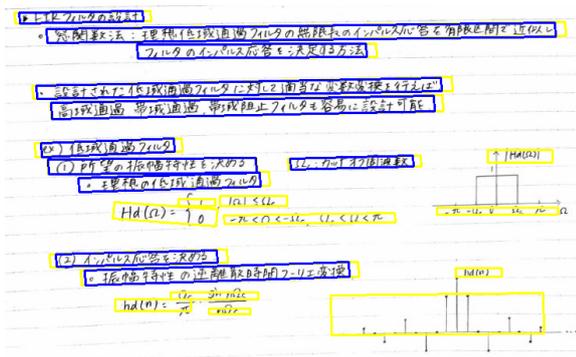


図5 最終出力テキスト領域(青)

U-Netによって抽出されたテキスト部分(図2青)とフィルタリングによって抽出された領域(図4(d))の組み合わせ

て,  $P_o, P_s$  で構成された画像  $B$  を用意する.  $\max$  フィルタは手書きノート画像内の輝度値が周辺と比較して最大の値を抽出されるため,  $B$  において,  $P_s$  に属する画素が抽出されにくく, 影領域を見逃しやすい. そこで,  $\max$  フィルタと同じフィルタサイズの  $\min$  フィルタを  $B$  に適用し, 影領域を適切に抽出する. このようにして, 得られた影付きの背景画像  $B'$  とする. 最後に, 元の画像から  $B'$  を減算し,  $[0, 255]$  で正規化する. 上記の一連の処理の例を図3に示す. 図3において, 影の影響が取り除かれていることが確認できる.

影領域が取り除かれた画像に対して, エッジ抽出を行い, テキストを行単位で抽出する. この際, 手書きノートには, グリッド線として下線が多いことを考え, Sobel フィルタを用いて垂直方向のエッジを抽出した2値化画像を得る. その後, 2値化画像に対して膨張および収縮操作(モルフォロジー変換)を実行する. モルフォロジー変換を行うことで, ノイズは抑制され, ノイズ以外のエッジ領域, すなわち, テキスト領域の画素が結合される. この際のモルフォロジー変換のフィルタサイズでは, 行の密着を防ぐため, 横方向のフィルタサイズよりも縦方向のフィルタサイズを小さく設定し, 実験的に  $40 \times 4$  とした. モルフォロジー変換を適用した結果を図4を示す. この方法で取得されたテキスト範囲は分類されていないため, より正確な結果を取



図6 2値化テキストと文字分割

得するには, U-Net によって生成された結果と組み合わせる必要がある. U-Net によって検出されたテキスト領域と上記の方法で得た2値化画像に対して AND 演算を行い, 各連結領域の外接矩形をテキスト範囲として, 文字分割を行う. 図5に示すように, 図や数式と区別して, 行単位でのテキスト領域の検出が可能である.

### 3.3 文字分割

まず, 前節で抽出された行について切り出した画像に対して, 大津2値化[8]を用いて2値化した後, 膨張処理を行う. 切り出した画像の2値画像  $I$  において, 各画素の座標を  $(x, y)$  としたとき, 対応する画素値  $I(x, y) = i, i \in \{0, 1\}$  とする. このとき,  $I$  の  $y$  軸(縦)方向に画素値を集計し, 連続して画素値が存在するときに文字が存在するとして, 文字分割を行う.

上記の文字分割を図6に示す. 上段に  $I$ , 中段に,  $y$  軸(縦)方向に画素値の集計値を示す棒グラフ  $h(x) = \sum_y I(x, y)$ , 下段に各スラッシュで分割された領域を表す.  $h(x)$  を左から右へスキャンし,  $h(x)$  の累計値が閾値を超えるまで加算していく. このスキャン時に閾値が満たされ,  $h(x) = 0$  となる列があれば, スキャンしたその部分までを一つ文字として分割し, 累計された画素数を0とする. このようにして  $h(x)$  の累計が閾値を満たしているかどうかを判断することで, わずかに隙間がある文字(例:小)を分割してしまわないようにできる. 閾値の設定は, 実験的に30を使用する.

表 1 U-Net による領域分割結果の評価

0:背景, 1:文字, 2:数式, 3:図, OUT:U-Net による推定値, GT:真値

	OUT = 0	OUT = 1	OUT = 2	OUT = 3
GT = 0	90.46%	6.90%	1.94%	0.71%
GT = 1	14.59%	80.57%	4.63%	0.2%
GT = 2	16.14%	13.5%	69.3%	1.05%
GT = 3	35.37%	4.76%	2.96%	56.91%

## 4. 実験

### 4.1 データセット

九州大学工学部の学部生を対象に開講されたデジタル信号処理の受講生 78 名に対して 11 回の講義の内容について手書きノートの写真を収集した。毎回各学生が任意のページ数の手書きノート画像を提出することができ、データセットには最大 10 ページの手書きノートを提出した受講者がいた。データセットに含まれる画像は A4 の紙のノートを撮影した手書きノート画像やタブレットやその他のツールを使用して手書きされた電子メモであり、それらの画像サイズとアスペクト比は均一でなく、画像の一部が鮮明でないものも存在する。手書きノートの領域分割の正解データは、背景、テキスト、数式、図のラベルで構成され、手でテキストと数式、図を矩形領域で囲みラベル付けすることで作成された。

デジタル信号処理の講義では、オンライン教育プラットフォーム Moodle とデジタル教科書閲覧システム BookRoll[9] を利用した。Moodle により、50 点満点の多肢選択問題 5 問を最終回にオンラインテストとして実施し、そのテストの点数を本実験の分析に利用する。なお、オンラインテスト時に教科書やインターネットを利用することは許可している。また、BookRoll により、受講者の教科書閲覧履歴が自動的に記録され、各受講者の教科書閲覧時間を分析に利用できる。ただし、電子教科書の閲覧ページが閉じられていない場合や閲覧履歴のデータベース保存時のエラー等の理由により、記録される時間が実際の閲覧時間よりも大幅に長くなることもある。

### 4.2 ページ内コンテンツの領域分割精度の評価

評価実験に用いたデータセットにおいて、テキスト領域は数式領域よりも割合が多く、図領域は相対的にさらに割合が少ない。データセットに含まれる各領域のバランスを担保するために、学習セットには第 3,5,11 回目の講義から 500 枚のノート画像を選択し、第 1,2 回目の講義の 160 枚のノート画像をテストセットに使用する。また、U-Net を学習する際のバッチサイズは 8 で、学習率は  $10^{-3}$  とし、最適化手法として Adam[10] を用いた。

U-Net による領域分割の評価結果を表 1 に示す。表 1 の

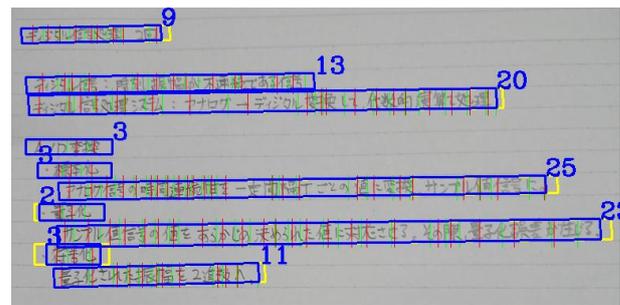


図 7 文字分割の例

OUT=1 から、文字領域の認識精度が高く、数式や図としての誤認識の割合が少ないことがわかる。また、数式領域の 13.5% がテキスト領域として誤認識されており、数式とテキストが隣接している部分では、誤認識が発生しやすいためである。正解データのラベル付けに矩形を使用しているので、図領域をアノテーションする際に、多くの空白部分が図領域としてアノテーションされる。そのため、図領域の 35% が背景として誤認識されると考えられる。どのページコンテンツの領域も 50% 以上で認識されており、本研究ではこの U-Net を手書きノートの定量化に利用する。また、図 7 に行単位の文字分割の結果例を示す。このようにして得られた文字分割により、文字数を各ページで計測したものを  $y$ 、実際の文字数を  $t$  としたときに、誤差率を  $\frac{|y-t|}{t}$  とする。評価に用いた画像に対して、誤差率を計算し、その平均は 18.82% であった。文字分割の精度は、ページの状況によって多少異なり、平均して 18.82% の誤差はあるが、本研究では、この文字分割の結果を利用してノート記述量を計算する。

### 4.3 手書きノートと学習活動の関係分析

学習活動データとして教科書閲覧時間、テストの点数と手書きノートの分析結果例を表 2 に示す。ノートのテキスト領域の定量化についてはテキスト領域から分割した文字数を用いる。また、ノート画像の解像度の影響を低減するために、ノートの図領域の定量化については図領域の画素数をページ全体の画素数で割った割合を用いる。表 2 の文字数得点および図得点は、各学生の文字数・図の画素数の割合を昇順で並び替えたランキングを意味する。したがって、最も多く文字・図を記述した学生の得点は高くなる。これら得点を加算することでノート総得点とした。本実験では、テキストや図が現れる初回授業のノートを分析対象とし、文字数・図の画素数の割合や教科書閲覧時間・テストの点数などを比較することで、ノートの記述量と学習活動の関連性を分析する。

まず、ノートの定量化に関する項目ごとに学習活動との相関係数を調べた。その結果、文字数とテスト点数の相関係数は 0.166、図の画素数の割合とテスト点数の相関係数は 0.2351 だった。また、文字数、図の画素数の割合と教科

表 2 ノートと学習活動のデータ例

学生	教科書利用時間 (秒)	テストの点数	図の画素数の割合	文字数	文字数得点 (ランキング)	図得点 (ランキング)	ノート総得点
A	22899	42.5	4.48%	404	43	49	92
B	18035	50	4.24%	645	59	46	105
C	43477	45	4.92%	122	12	51	63
D	14219	50	0%	763	63	1	64
E	43095	42.5	2.65%	198	25	38	63
F	18618	35	6.11%	337	39	55	94
G	6883	32.5	0.32%	105	6	22	28

書閲覧時間の相関係数は、0.2146 と 0.408 だった。

次に、文字数と図の画素数の割合を組み合わせたノート総得点と学習活動との相関係数を調べた。前述のとおり、ノート総得点を計算するために、文字数と図の画素数の割合のランキングの値を計算に用いた。その結果、テストの点数および教科書閲覧時間の相関係数は、0.2524 と 0.4337 だった。文字数と図の画素数の割合を組み合わせることで、ノートの記述量の評価と各学習活動との相関が強まることが確認できた。

図 8, 9, 10 に教科書閲覧時間、テストの点数、ノート総得点に関する散布図を示す。図 8 に着目すると、教科書閲覧時間が長くなると、テストの点数が相対的に低い 30 点未満の学生数が減ることがわかる。

図 9 は、ノートの記述量と学生の教科書閲覧時間の関係を表す。ノートの記述量と教科書閲覧時間は、先述の通り、正の相関を示していることがわかる。

図 10 は、テストの点数とノートの記述量の関連性を表す散布図である。今回の講義でのテストでは、受験中に教科書やインターネットを利用することができたため、学生は高得点を獲得したと考えられる。しかし、ノートの記述量が少ない学生に低いテストの点数が発生しやすいことが確認でき、ノート総得点が 80 未満のときにテストの点数が約 30 点以下の学生に偏りが見られる。

ここまでの分析結果から、ノートの記述量と学習時間やテストの点数には弱相関があり、ノートの記述量が多い学生は学習時間が長く、テストの点数が高い傾向にあることがわかる。この結果から、教育システムによるノートの自動評価により、ノートの記述量と学生の学習活動を分析して、教師は手書きノートの閲覧の負担なく学生の学習状況を把握できる可能性がある。また、教科書の閲覧ログを収集するシステムを利用していなくとも、ノートの記述量による学生の教科書閲覧時間の推定が可能であることも示唆された。

## 5. おわりに

本研究では、ページセグメンテーションによって手書きノートに含まれるテキストの文字数や図の量を定量化し、定量化されたデータと学生の学習活動データの関係を分析した。提案手法ではニューラルネットワークによる領域分

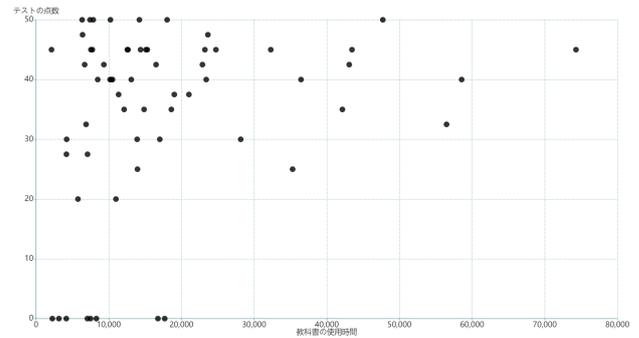


図 8 ノートを提出した 63 人の学生のテストの点数と教科書閲覧時間

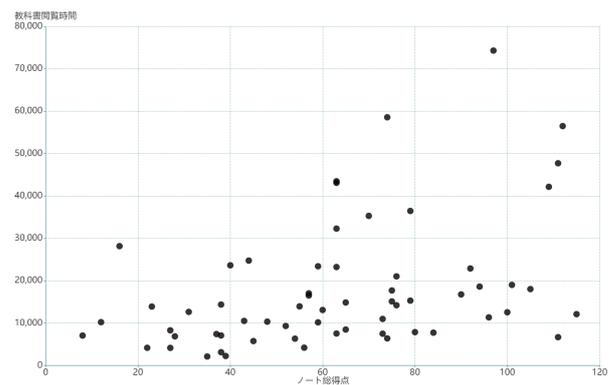


図 9 ノート総得点と教科書閲覧時間

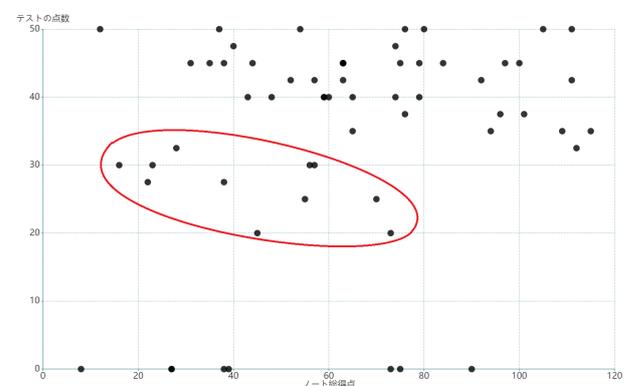


図 10 テストの点数とノート総得点

割手法と画像処理手法を用いることで、ページ内の異なる種類の各領域を分割して、ページ内の文字数を計測した。

また、分析において、文字数と図の画素数の割合を組み合わせることで、手書きノートと学生の学習活動の間の弱相関関係を示した。

しかし、ノートの評価方法については、文字や数式、図などを様々な種類の領域を統合的に評価する方法について、さらなる研究が必要である。今後は、ページセグメンテーションの精度を高めるとともに、文字分割の技術の精度向上を図っていく。

## 謝辞

本研究は、JST AIP Grant Number JPMJCR19U1 および JSPS 科研費 JP18H04125 の助成を受けたものです。

## 参考文献

- [1] 岸 俊行, 塚田裕恵, 野嶋栄一郎: ノートテイキングの有無と事後テストの得点との関連分析, 日本教育工学会論文誌, Vol. 28, No. suppl, pp. 265–268 (オンライン), DOI: 10.15077/jjet.KJ00003730718 (2005).
- [2] 大倉孝昭ほか: 手書きレポートを対象とした学習分析支援システム, 研究報告教育学習支援情報システム (CLE), Vol. 2018, No. 2, pp. 1–4 (2018).
- [3] 渡邊栄治, 尾関孝史, 小濱剛ほか: 講義や自習における学習者の手書きノートの分析 (第 2 報), 研究報告ドキュメントコミュニケーション (DC), Vol. 2018, No. 1, pp. 1–8 (2018).
- [4] 渡邊栄治, 貝原準, 川島健嗣ほか: 手書きノートの分析に関する研究: 第 1 報, 甲南大学紀要. 知能情報学編, Vol. 1, No. 2, pp. 197–208 (2008).
- [5] Jia, W., Zhong, Z., Sun, L. and Huo, Q.: A CNN-based approach to detecting text from images of whiteboards and handwritten notes, *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, IEEE, pp. 1–6 (2018).
- [6] He, D., Cohen, S., Price, B., Kifer, D. and Giles, C. L.: Multi-scale multi-task fcn for semantic page segmentation and table detection, *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 1, IEEE, pp. 254–261 (2017).
- [7] Ronneberger, O., Fischer, P. and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241 (2015).
- [8] Otsu, N.: A Threshold Selection Method from Gray-Level Histograms, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, pp. 62–66 (online), DOI: 10.1109/TSMC.1979.4310076 (1979).
- [9] Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K. and Yamada, M.: E-Book-based learning analytics in university education, *International conference on computer in education (ICCE 2015)*, pp. 401–406 (2015).
- [10] Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).