

遠隔教育における自立した学習を支援するための 教材のページ認識

今井真宏^{†1} 荒井正之^{†1}

教材を使って自学自習する学習者は、理解できない課題をそのまま放置してモチベーションが下がり、学習を諦めることが多く、今も教育現場の課題の1つである。従来の学習システムに多く見られる科目や単元を選択して講義動画を閲覧する方法では、個々人の理解できない問題にフィットせず、解決にいたらない。そこでユーザが教材のページをスマートフォンで撮影し送信すると画像認識を用いてページの問題を特定し、必要なコンテンツをすぐに返答するシステムを提案する。画像認識には、Bug Of Featuresを用いた特定物体認識と、画像の姿勢推定の情報を用いて機械学習した識別器を合わせる手法を用いる。教材 100 冊 11,935 ページを分類対象とした実験を行い、認識率 94.7%と処理時間約 2.7 秒を得ることができた。

Textbook Page Recognition for self-study in distance learning

MASAHIRO IMAI^{†1} MASAYUKI ARAI^{†1}

The learners who do self-study by using Textbooks are likely to leave the problems they can't understand and often give up studying since the incomprehensibility leads to a decrease in motivation. This is still a problem in the field of education. In existing common learning system, the learners choose subjects or course units and watch videos of the lectures. This method doesn't offer a solution for the problems which each individual learner has. We therefore propose a new system. When users take and send a picture of the page of study materials, this system provides useful contents immediately by identifying the problem in the page using image recognition. For the image recognition, we use the specific object recognition of "Bug Of Features" and a classifier performed machine learning by using information from posture estimation of the picture. We conducted an examination on 11,935 pages from 100 Textbooks and achieved 94.7% recognition rate and about 2.7 seconds of processing time.

1. はじめに

近年、IT を使った遠隔教育サービスが注目を集めており、欧米を中心に Udemy[1]、Coursera[2]や NovoED[3]などの遠隔教育サービスが人気を博している。ハイレベルで実践的な講義を無料で受講することができるため、学習環境を求めるユーザが利用して自学自習を進めることができる。このような教育サービスの勢いは世界的に増しており、日本でも schoo (スクー)[4]をはじめとしたサービスが展開されている。しかしモチベーションの高い学習者がいる一方で、数多くのモチベーションが低い学生に継続して学習してもらうことは、教育サービスの大きな課題である。学習を諦めるきっかけは人によって異なるであろうが、自分 1 人では理解できない課題に遭遇したとき、そのまま放置してしまい、学習そのものを諦めてしまうことが多いと考えている。

本研究は、ユーザが学習を諦めるきっかけを減らし学習を継続できるようにするために、画像認識技術を用いたシステムを開発することを目的としている。インターネットさえあれば、塾(インターネット上の仮想的な塾を除く)などの特別な環境を必要としない、自分で学習すること(自立した学習と呼ぶ)をサポートするシステムである。

今までの自学自習では YouTube[5] などにアップロードされた様々な教育動画が利用されることが多く、これらは

基礎的または汎用的な内容を説明するものであるためユーザ自身が抱える問題・課題に対しての「応用」が難しいことが問題となっている。これを解決するためには、提案システムが提供する、「ユーザが必要としたときに 1 人 1 人の問題・課題にフィットしたコンテンツをすぐに得られる機能」(提案システムにおけるコンテンツとは、ユーザの学習をサポートする動画や画像などを示す。)が重要である。以下では、システムの機能と実装方法について述べる。

2. システムの要求仕様と概要

本システムのユーザは、主に問題集や教科書などの書籍で自学学習を進める学生を想定している。概して彼らは、途中で自分 1 人では解決できない問題・課題に遭遇する。その時、長時間理解できない・先に進めない状態を続けると学習をギブアップする可能性がある。(図 1)

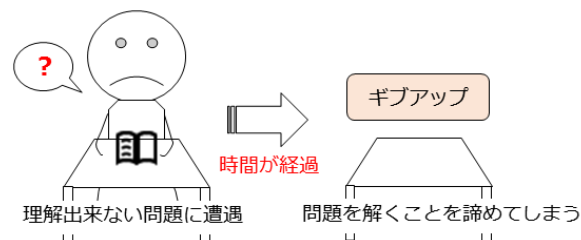


図 1 本システムが対象とする学習者(理解できない問題に遭遇し長い時間経過するとギブアップしてしまう)

^{†1} 帝京大学大学院理工学研究科
Graduate School of Science & Engineering, Teikyo University

提案システムを用いることにより、図2に示すような良好な学習サイクルを構築できると考えている。

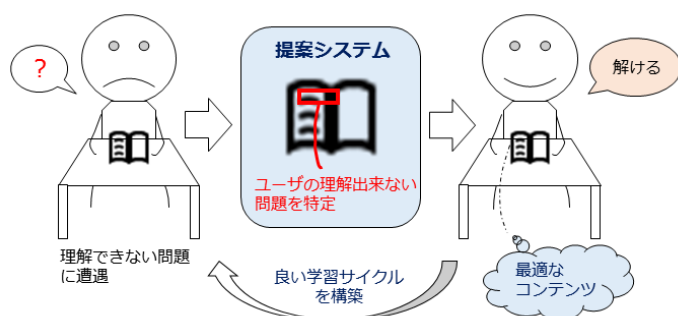


図2 提案システムを利用した良好な学習サイクル

理解できない問題に遭遇した時に提案システムを利用すれば、すぐに問題に合った最適なコンテンツを得ることができる。問題をすぐに解決して先に進むことで学習を円滑に進めることが可能である。

本システムには次のような仕様が求められる。

- 1) ユーザへすぐにレスポンスする
- 2) ユーザの理解できない問題・課題を特定する
- 3) 誤ったコンテンツをレスポンスしない

要件1)を満たすためにはシステムが自動で判定して、すぐにレスポンス可能な入力データが必要である。また、要件2)を解決する方法として、「入力データの意味を解析する方法」と「入力データの場所（書籍のページと位置）を認識する方法」が考えられる。「入力データの意味を解析する方法」では、文字認識(OCR)や図形の認識の後、それらで表される問題・課題の意味表現と知識による書き換えを経て回答を計算するが、いくつもの認識・解析・推定処理を組み合わせたシステムが必要である。国立情報学研究所における取り組み[6][7][8]が知られているが、問題の種類に問わず迅速なレスポンスを返すことは難しい。よって、本研究では「書籍のページと位置を認識する方法」を選択する。

認識に必要な入力データの候補について、適用可能な教材と環境に対する適正を比較する。画像、バーコードなどのマーカー、音声、タイピングによる文字入力 of 4つを入力候補として、教材の変化や使用環境に対する適用可能性を検討した結果を表1に示す。

表1より、画像データが最も多くの教材の改定・言語の変更や学習者の使用環境に適用可能であり、優れていることがわかる。よって、入力データには画像を採用する。

要件3)は学習を支援するシステムとして重要である。ユーザが問題・課題に対して異なるコンテンツを閲覧すると、誤った考え方や方法を学習する可能性があるためである。提案システムでは要件3)を満たすため、教材のページを正しく特定できたかどうかを判定し、誤分類の検出処理を行う。

画像データを用いた本システムの全体の流れを図3に示す。

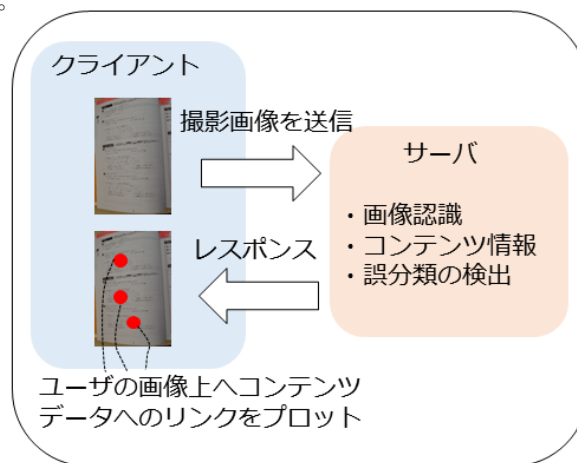


図3 システム全体の流れ

ユーザは本システムのクライアントサイドで撮影した画像をサーバに送信する。サーバは画像認識によって問題のページとユーザの画像上の問題の位置を特定した後、動画な

表1 入力データの違いによる適用可能な教材と学習者の使用環境

		画像	マーカー	音声	文字入力
多様な教材への適用	テキスト名や問題番号がない、プリントやコピーした教材・問題	○	○	△	○
	外国語や難しい漢字を使用しているなどの理由で音読ができない教材・問題	○	○	×	×
	出版済みの教材・問題	○	×	○	○
	文字情報を含まない教材・問題	○	○	×	×
	教材の変更・改定	○	×	△	○
	教材を記述する言語の変更	○	○	○	○
学習者の使用環境	入力時間がとれない場合	○	○	△	×
	図書館など音を出すのが憚られる場所	○	△	×	○
	電車の中などカメラが利用できない場所	×	○	○	○

※記号の意味 ○：適用可能、△：適用にコストがかかる、×：適用不可能

どのコンテンツの閲覧に必要な情報を生成し、クライアントサイドへレスポンスする。クライアントサイドでは、受け取ったデータを元にユーザの撮影した画像上に目印をプロットする。ユーザはプロットされた目印のうち、目的の問題に対応するものを選択することで、コンテンツを得ることができる。

次に、提案システムの画像認識の手法について述べる。

3. システムに用いた画像認識技術と処理の流れ

書籍のページには、固有のフォーマットやレイアウトの形状が特徴的に描かれるためことが多い。一方、出版社やシリーズが同一の場合、フォーマットやレイアウトが同一もしくは似ているページが多数含まれるため、各ページの中で特徴的な形状が必ずしも分類に役立つとは限らない。そのため大量の局所的な特徴(local feature)を抽出する BoF (Bug Of Features) の手法を用いて特定物体認識(Specific Object Recognition) を行い、書籍のページを認識する。

3.1 局所特徴量を使った特定物体認識

BoF を用いた特定物体認識の処理の流れを図 4 に示す。

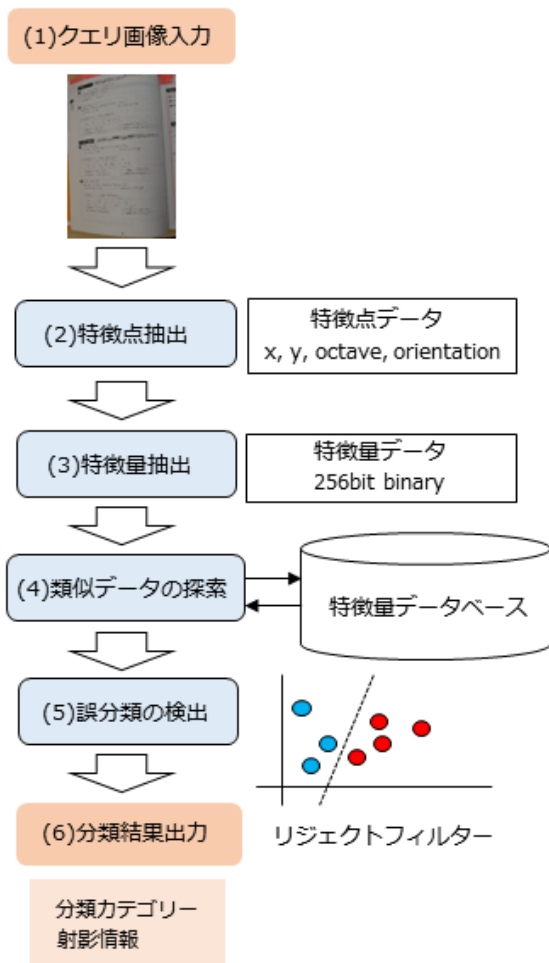


図 4 システムで用いた特定物体認識処理の流れ

図 4(1)のクエリ画像入力とは本システムへの入力データであり、ユーザが送信する画像である。図 4(2)は画像から特徴点の採取する処理であり、その後特徴点の周りから図 4(3)で示すように特徴量を抽出する。続けて図 4(4)では特徴量データベースを探索し、分類カテゴリーの候補を検索する。特徴量データベースは事前に識別対象の複数の画像から特徴点および特徴量を採取して作成する。図 4(5)の誤識別と考えられる事例を排除する処理を経て、図 4(6)の物体認識の出力を得る。出力は画像の分類カテゴリーと、クエリとデータベースの画像との特徴点の対応関係を表す射影情報である。

3.2 特徴点・特徴量の抽出

図 4(2)の特徴点抽出及び図 4(3)の特徴量抽出処理には、特徴点抽出及び特徴量同士の距離計算が高速なバイナリ特徴量 ORB [9] を用いる。特徴点の抽出に使用する FAST [10] はスケールに対してロバストではないため、画像の S 階層のピラミッド(図 5)を生成して特徴点候補を抽出し、Harris の尺度[11]を用いてレスポンスの良いコーナーを選択する。

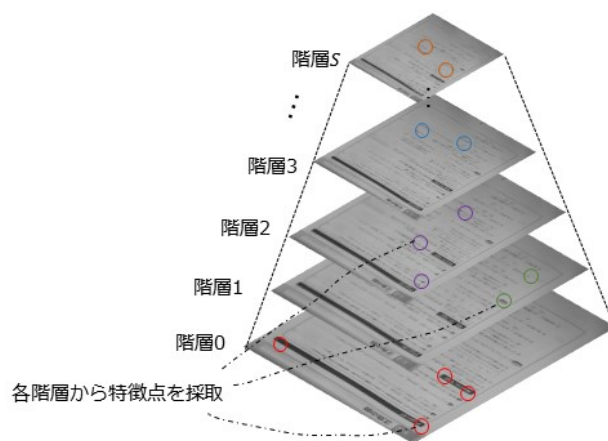


図 5 画像ピラミッド

特徴点の方向は、Intensity Centroid[12]を用いて決定する。式(1)から(4)を用いて 0 次モーメント m_{00} 、1 次モーメント m_{10} 、 m_{01} 及び重心 C を得る。 C の方向を式(5)によって求

$$m_{00} = \sum_{x,y} I(x, y) \quad (1)$$

めて、特徴点の方向 φ とする。 $I(x, y)$ は点 (x, y) の輝度を表す。

$$m_{10} = \sum_{x,y} [x I(x, y)] \quad (2)$$

$$m_{01} = \sum_{x,y} [y I(x, y)] \quad (3)$$

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (4)$$

$$\varphi = \tan^{-1} \left(\frac{m_{01}}{m_{10}} \right) \quad (5)$$

ORB の特徴量は rBrief [9]を用いる。rBrief は、Brief [13] に回転に対する不変性を補った 256bit のバイナリ特徴量である。図 6 は、本システムで使った 256 組のバイナリテストを回転なしで可視化したものである。バイナリテストには、赤線で結ばれた 2 点の間の輝度の大小関係を使用する。

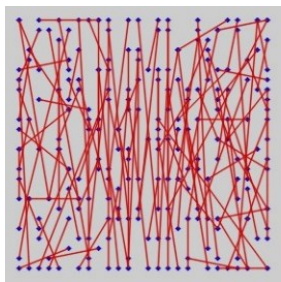


図 6 本システムで用いたバイナリテストの組

3.3 類似した特徴量の探索

図 4(4)の類似データの探索を高速化するため、図 7 に示すようにハミング空間における LSH(Locality Sensitive Hashing) [14]を用いる。

特徴点 x_i 、 $i \in \{1 \dots N\}$ から抽出した特徴量を v_i とし、 v_i から M ビットのハッシュ値を生成する関数を h_j 、 $j \in \{1 \dots L\}$ とする。

LSH は識別対象の画像から抽出した特徴量を用いて計算したハッシュ $h_j(v_i)$ をキーとした、 $h_j(v_i)$ とバケットの組をもつ L 個のハッシュテーブルを事前に作成する。入力画像から抽出した特徴量と類似したベクトルを探索する際に、ハッシュキーが一致したバケット内のみを探索することによって処理時間を短縮する。

さらに、メモリ使用量の削減のため、multi-probe LSH [15][16]を用いる。multi-probe LSH は M のうち t ビットが異なることを許容して検索するバケットを選択することでテーブル数 L の値を小さくし、RAM に展開するデータを節約できる。 v_i とバケットの中にある類似候補がもつ特徴量とのハミング距離を計算し、昇順に第 k 位までソートして選択することで、テーブルごとに入力画像中の特徴点 x_i とデ

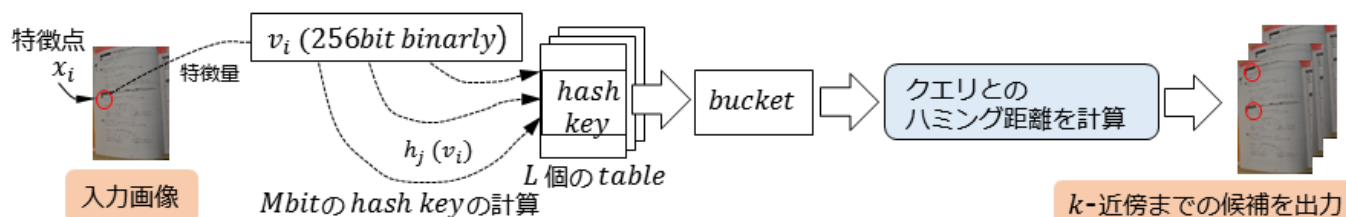


図 7 Locality Sensitive Hashing による類似特徴量の探索

ータベース上の特徴点との類似候補 $\{c_{i1}, \dots, c_{ik}\}$ を得る。(図 8) ただし、 x_i に対して $M - t$ ビット以上一致した候補が見つからないこともあるため、最大で k 個の候補を得る。

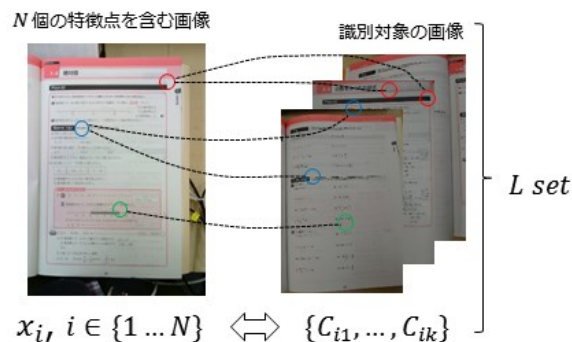


図 8 N 個の特徴点をもつ入力画像と識別対象の画像中の特徴点の対応

3.4 誤分類の検出

3.3 の処理によって得た類似候補からインライアを抽出するため、RANSAC [17]を用いる。クエリ画像中の特徴点数 N 、LSH のテーブル数を L とし、 k 近傍までを候補とすると N 個の特徴点に対する類似候補の最大数は式(6)で表せる。

$$\max C_n = L * N * k \quad (6)$$

$\max C_n$ 個の点の中からランダムに選択した 4 組の点 (X, Y) を使って生成したホモグラフィ行列 H によって入力画像上の特徴点 x_i 、 $i \in \{1 \dots N\}$ を識別対象の画像上へ射影した点を $y_i = H(x_i)$ とする。 y_i を中心とした半径 ε の円の円周もしくは内側に識別対象の画像から抽出した特徴点をただ 1 つ含む場合、インライアと判断する。以上の処理を繰り返す行い、 N のうちインライアの割合 w が最も大きい H_{max} を入力画像から識別画像への射影行列として採用する。繰り返しの回数 R は、ランダムに選択した 4 点がインライアである確率を u とし、繰り返しの結果全てがインライアである確率を p とすると式(7)で表せる。

$$R = \frac{\log(1-p)}{\log(1-u^4)} \quad (7)$$

次にインライアの割合 w をスコアとして分類候補を順位付けする。しかし、インライアの割合 w は画像から抽出可能な特徴点数や画像中に含まれる物体の大きさ、同一フォーマットやレイアウトが文書に含まれる割合などに依存す

るため、正分類と判定するためのスコアの閾値 $R, R \in \mathbb{R}$ を用いても分類結果が誤っているかどうかの判定が難しい。図 9 に分類対象外のページを送信した誤分類事例(緑の点)とページの分類に成功した正分類事例(青の点)のスコアをサンプリングしたものを示す。スコアの閾値の判定では、緑でプロットした多くの事例を正分類と判定する可能性がある。

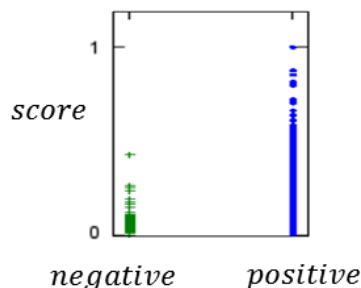


図 9 誤分類事例(緑)と正分類事例(青)のサンプリング

そこで本システムでは H_{max} を使って入力画像上の矩形を識別対象の画像上へ射影することで得られる 4 点の幾何学的特徴を利用して誤分類を判定する。射影後の 4 点でできる四角形を R としたとき、 R の周囲、内角、隣り合わない辺同士の関係、4 点の共分散を用いる。図 10 はスコアと R の周囲、隣り合わない辺どうしの内積をサンプリングしたものである。正事例を青の点、誤事例を緑の点で表す。

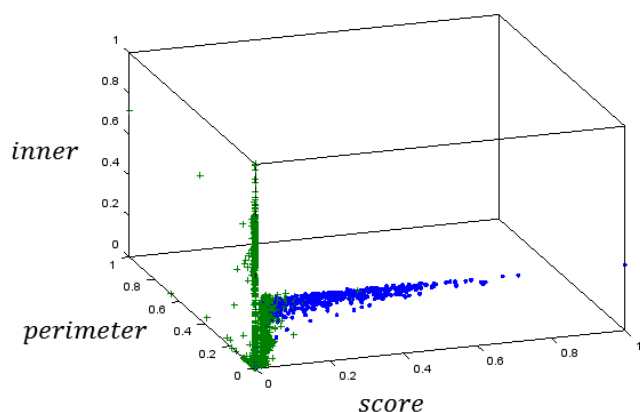


図 10 スコアと R の周囲、隣り合わない辺どうしの内積のサンプリング

このようにして、正事例と誤事例の新たな識別境界が得られる特徴 X を用いて、ロジスティック回帰により機械学習を行い、正事例を判定する識別器を生成する。

識別器はパラメータ θ と特徴量 X の積によって得られる。パラメータ θ を学習するためのコスト関数は式(8)、(9)、(10)を用いた。

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m (-y^{(i)} \log(h) - (1 - y^{(i)}) \log(1 - h)) \quad (8)$$

$$h = h_{\theta}(X^{(i)}) = g(\theta^T X) \quad (9)$$

$$g(z) = \frac{1}{1 + e^{-z}} \quad (10)$$

式(11)を満たした時、正分類と判定して分類結果と H_{max} を出力する。 α は定数である。

$$\theta^T X \geq \alpha \quad (11)$$

4. 評価実験

4.1 実験に用いたデータセット

問題集や学習参考書などの教材のページを対象としたシステムの分類器を評価するため、分類対象の画像として、数学の教材 100 種類の合計 11,935 ページを用意した。ユーザが送信することを想定したクエリ画像には、分類対象に含まれる教材のページをカメラの姿勢やカメラからの距離、照明を変化させて撮影した 1,540 枚を使用する。なお、1,540 枚は全て異なる教材の異なるページである。訓練画像の解像度は 1128x750 ピクセル、クエリ画像の解像度はスマートフォンに搭載される多くのカメラで撮影が可能な 640x480 ピクセルを用いた。

分類器の精度計測に用いた PC は、Intel Core i7 3.4 GHz processor(6 コア 12 スレッド) CPU、16GB RAM である。

4.2 教材画像からの特徴点抽出

図 12 は、最大特徴点数毎に、画像から得られた特徴点を緑の点でプロットし、教材のレイアウト周辺の領域と数字周辺の領域をそれぞれ切り出したものである。

図 12 の A、B、C、D はそれぞれ最大 500、1000、1500、2000 点を抽出してプロットした結果である。図 12 の A か



図 11 最大特徴点数による教材のレイアウト、数字周辺の特徴点比較

らわかるように最大で 500 点を抽出した場合、教材のレイアウトやフォーマットからは多くの特徴点を抽出できるが、数字や演算子からは抽出できていないことがわかる。一方、図 12 の D に示すように、2000 点以上抽出することで数字や演算子からも多くの特徴が得られ、数式が多く存在する教材のページに有効であることがわかる。

4.3 特定物体認識の精度と処理速度

LSH はテーブル毎に並列で処理を行うことが可能であるため、12 スレッドを使って並列に探索処理を行う。

表 2 は、basic な LSH を用いた場合の、特徴点数ごとの認識精度と処理時間の比較である。テーブル数 $L = 16$ 、ハッシュキーのビット数 $M = 26$ とした。精度は、システムがクエリ画像を正しい教材のページに分類した割合である。

表 2 basic な LSH を用いた場合の精度と処理時間

最大特徴点数	精度(%)	平均処理時間(msec)
500	69.22	445.6
1000	85.77	669.5
2000	93.96	1124.5
4000	96.23	1992.1

表 3 は、multi-probe LSH を用いて、誤差として許容するビット数を 1、テーブル数 $L = 4$ とした場合の精度と処理時間の比較である。テーブル数を 1/4 にすることで、LSH のテーブルの展開に必要な RAM 使用量を basic な LSH の約 25% に節約し、精度を向上することができる。処理時間は 2683.4(msec) となるが、学習システムとして許容できる範囲だと考えている。

表 3 multi-probe LSH を用いた場合の精度と処理時間

最大特徴点数	精度(%)	平均処理時間(msec)
500	85.25	643.8
1000	93.37	938.6
2000	95.25	1484.7
4000	97.72	2683.4

次に、クエリ画像の階層数とスケールについて検討した結果を図 12 に示す。

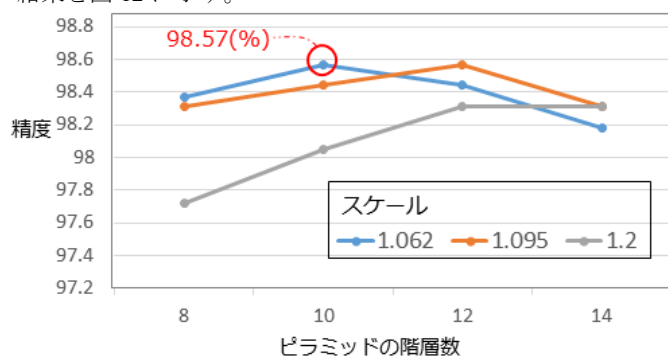


図 12 クエリ画像ピラミッドの階層とスケールの変化による精度比較

ピラミッドの階層数 10、スケール 1.062 とすることで、精度 98.57(%)を得ることができた。データベースのサイズを増やさずに精度を向上できることが確認できた。

4.4 誤分類の検出精度

誤分類の検出に用いるリジェクトフィルターは 4.1 のデータセットに含まれるクエリ用画像のうち半分に、システムで得られたユーザからのクエリ画像を追加したデータを使用する。学習にはロジスティック回帰を用い、学習に使わなかった残りのデータを精度測定に用いた。精度には式 (12) の $F - Measure$ を用いる。

$$F - Measure = \frac{2 \text{ Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (12)$$

表 4 は、リジェクトフィルターの学習に、3.4 で述べたインライアの割合によるスコアのみを用いた場合と射影情報によって生成した特徴量を用いた場合との精度の比較である。提案システムで用いた方法により、スコアのみを用いた方法よりも精度が向上していることがわかる。

表 4 学習に使用した特徴量によるリジェクトフィルターの精度比較

学習に使用した特徴量	精度(%)
インライアの割合によるスコア	89.66
射影情報によって生成した特徴量	96.09

4.5 評価実験の結果

提案システムの精度と処理時間を表 5 に示す。分類器全体の認識精度は 94.71(%)、処理時間 2683.4(msec)を得ることができ、実用に耐えうる性能を得られたと考えている。

表 5 分類器の精度と処理時間

分類器の種類	精度(%)	処理時間(msec)
(a)物体認識のみ	98.57	2683.4
(b)誤分類の検出のみ	96.09	
(a)*(b)分類器全体	94.71	

5. おわりに

本研究では、主に遠隔教育において、学習者が学習を諦めてしまう原因の 1 つである、自分自身で理解できない問題を放置してしまう問題を解決するために、画像認識を用いたシステムを提案した。教材の画像からページを特定することで、すぐに必要な解説を得られるシステムである。教材のページの認識では、特徴量に ORB[9]を使った Bug Of Features の手法による特定物体認識と、ロジスティック回帰により学習した画像の分類の正しさを判定する識別器を使用した方法で、分類精度 94.70%、平均処理時間 2683.4(msec)を得ることができた。特に本システムのように

特定物体認識に正しい分類結果が求められる場合、物体認識で得られた特徴を利用した誤分類の識別器を組み合わせる手法は有効である。

今回の実験において誤分類となった多くのページに共通する特徴は、2つか3つの項しかもたない多項式の計算問題のみが、同じ配置で記載されていることである。各多項式の図形的な相違点は主に係数部分の数字である。そのため今後の課題は、抽出した特徴量から、より視覚的に類似した教材のページを分類するために重要な情報を学習することである。また、分類に重要な領域から特徴の共起を学習する方法なども今後の検討課題である。

参考文献

- 1) Udemy <https://www.udemy.com> (アクセス 2014/10/21)
- 2) Coursera <https://www.coursera.org> (アクセス 2014/10/21)
- 3) NovoED <https://noved.com> (アクセス 2014/10/21)
- 4) schoo(スクー) <http://schoo.jp> (アクセス 2014/10/21)
- 5) YouTube <https://www.youtube.com> (アクセス 2014/10/21)
- 6) 松崎拓也, 岩根秀直, 穴井宏和, 相澤彰子, 新井紀子. 深い言語理解と数式処理の接合による入試数学問題解答システム 人口知能学会全国大会論文集(CD-ROM) 27th, ROMBUNNO. 2A4-1, 2013
- 7) 岩根秀直, 松崎拓也, 穴井宏和, 新井紀子. 数式処理による入試数学問題の解法と言語処理との接合における課題. 人口知能学会全国大会論文集(CD-ROM), 27th, ROMBUNNO. 2A4-2, 2013
- 8) T.Matsuzaki, H.Iwane, H.Anai and N.Arai, The Complexity of Math Problems -- Linguistic, or Computational? Proceedings of the 6th International Joint Conference on Natural Language Processing, pp73-81 2013
- 9) E.Rublee, V.Rabaud, K.Konolige, G.Bradski. ORB: an efficient alternative to SIFT or SURF. In Proc. of IEEE International Conference on Computer Vision, 2011.
- 10) E.Rosten, T.Drummond. Machine learning for high-speed corner detection. Proc.of European Conference on Computer Vision, pp. 430-443, 2006.
- 11) C.Harris, M.Stephens. A Combined Corner and Edge Detector, Proceedings of The 4th AlveyVision Conference (Manchester, UK), pp. 147-151, 1988.
- 12) L.Rosin. Measuring Corner Properties. Proc of Computer Vision and Image Understanding, Vol.73, No.2, pp.291-307, 1999.
- 13) M.Calonder, V.Lepetit, C.Strecha, P.Fua. Brief: Binary Robust Independent Elementary Features. In European Conference on Computer Vision, 2010.
- 14) A.Gionis, P.Indyk, R.Motwani. Similarity Search in High dimensions via hashing. In M.P.ATkinson, M.E.Orlowska, P.Valduriez, S.B.Zdonik, and M.L.Brodie, editors, VLDB'99, Proceedings of 25th International Conference on Very Large Data Bases, September -10, 1999, Edinburgh, Scotland, UK, pp 518-529. Morgan Kaufmann, 1999.
- 15) Q.Lv, W.Josephson, Z.Wang, M.Charikar, K.Li. Multi-probe LSH: efficient indexing for highdimensional similarity search, Proceedings of the 33rd international conference on Very large data bases, pp.950-961, VLDB '07, VLDB Endowment, 2007.
- 16) M.Slaney, Y.Lifshits, J.He. Optimal Parameters for Locality-Sensitive Hashing. Proceedings of the IEEE Vol.100, No.9, pp 2604-2623, September 2012.
- 17) M.A.Fischler, R.C.Bolles. Random Sample Consensus: a Paradigm For Model Fitting With Applications To Image Analysis And Automated Cartography. Commun. ACM, vol.24, no.6, pp.281-395, June 1981.