

Web上の画像から意味のある式を 抽出するための数式フィルタの作成

山田 奉子^{1,a)} 村上 晴美¹

概要: 数学情報検索 (MIR) に関する研究は多数あるが, MathML や TeX で表された数式を解析するものがほとんどである. 一方で, Web 上には数式画像を含む文書が多数存在するが, 数式画像に関する研究は少ない. 本研究は, Web 上の HTML 文書から, $\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K(\frac{x-x_i}{h})$ のような「コンパクトに表現された単独で意味のある数式」を抽出するものである. 文書中の画像を教師あり機械学習の手法で, 意味のある数式とそうでないものに二値分類する. 深層学習以外の機械学習の手法と畳み込みネットワーク (CNN) の手法を比較した結果, VGG16 を用いた転移学習とファインチューニングを行った CNN の手法が一番良く, マッシュューズ相関係数 (MCC) が 0.771, F 値が 0.814 であった. さらに応用事例として, 数学用語をクエリとして Web 検索を行い, その用語と関連する数式を上位 10 件提示するという先行研究に対して, 先行研究で使用した SVM による分類器を本研究で最も性能が良い CNN に置き換えたところ, F 値・MAP・MRR のいずれも先行研究を上回った.

1. はじめに

近年, [1] や [2] で述べられているように, 数学情報検索 (mathematical information retrieval: MIR) の重要性が認識され, 多くの研究がなされてきた. また, 国立情報学研究所による NTICIR でも, パイロットタスクを含め過去 3 回 math タスク^{*1}が行われている [3]. しかし, その多くは, MathML や TeX で表現された数式を解析することによって, 類似数式を検索するものである [4][5]. Web 上には数式画像を含む HTML 文書 (以下, Web 数式文書) が数多く存在するが, 数式画像に関する研究は少ない.

Web 数式文書中の数式画像を得たければ, その文書をダウンロードし作成されたフォルダ内を見ればよい. しかし, 数式以外のグラフやアイコンなどもあり, しかも数式に使われるフォントは x のように通常とは異なるため画像で表されることが多く, フォルダ内のファイル数は多くなる^{*2}. x や γ , または公式等を導出する長大な数式ではなく, 例えば「カーネル密度推定量」を表す

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (1)$$

¹ 大阪市立大学
Osaka City University, 3-3-138, Sugimoyo, Sumiyoshi, Osaka
558-8585, Japan

^{a)} d18ud512@eb.osaka-cu.ac.jp

^{*1} <http://ntcir-math.nii.ac.jp/>

^{*2} 「erf 関数」をクエリとして Web 検索を行ったところ, 上位 1 件目のダウンロードフォルダの画像数は 91 個であった.

のような数式を, それらの中から探すのは時間がかかる.

本研究は, Web 数式文書中にある多数の画像から, 式 (1) のような「コンパクトに表現された単独で意味のある数式」を抽出するものである. 以下, それらの数式を「意味のある数式」と呼ぶことにし, 意味のある数式とその他の画像を教師あり機械学習の方法で二値に分類する方法を提案する.

2 章で関連研究について, 3 章で提案手法について, 4 章で, まず深層学習以外の機械学習を用いた分類について, 次に深層学習を用いた分類について述べる. その後, 5 章で, 応用事例について実験をした後, 6 章で議論しまとめる.

2. 関連研究

数式検索とは, 「クエリまたは対象文書に数式を含む検索 [1]」のことであり, 例えば, 数式中の部分数式に着目した [6], [7] や, 数式の構造とそれらに結びついたテキスト文書を用いて数学文書を分類する [8] がある. また, 狭義には, 「与えられた数式に対して類似構造や意味を持つものを検索すること [2]」であり, [4] と [5] がそれにあたる.

クエリに数式画像を用いたものには, [9] や [10] がある. [9] は記号を数式画像から取り出し, 高さ, 幅, 垂直位置により表現して, あらかじめ収集したデータセット内の数式との類似度を算出している. [10] は, 数式画像を入力文書から抽出後, その画像の特徴量を抽出し, それらの値を用いて類似するものを画像データベースから出力している.

(1) $\hat{f}_h(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x-x_i}{h}\right)$	(5) $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$
(2) P	$= \int_{-\infty}^0 x^2 f(x) dx + \int_0^{\infty} x^2 f(x) dx$
(3) \log	$= -\int_{-\infty}^0 t^2 f(-t) dt + \int_0^{\infty} x^2 f(x) dx$
(4) \leq	$= 2 \int_0^{\infty} x^2 f(x) dx$
	$= 2k^{-1} \left[B\left(\frac{k}{2}, \frac{1}{2}\right) \right]^{-1} \int_0^{\infty} x^2 \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} dx$

図 1 数式例

本研究は特定のデータベースからの検索ではなく一般の Web 数式文書を対象にしたものであり、数学文書内の数式の中から意味のある数式を抽出するものである。

3. 提案手法

3.1 数式

数式には様々な形のものがある。例えば、図 1 中の (1) は「カーネル密度推定量」を表す意味のある数式であるが、(2) は文脈により数式とそうでない場合とがある*3。(3) と (4) は数式ではない。(5) は t 分布の期待値および分散を求める途中の計算式である。(1) のような「コンパクトに表現された単独で意味のある数式」がどのような特徴を持つかを後述する実験から決定する。

3.2 数式の色情報

Web 上の数式画像は、見かけ上白黒二値に見えることが多いが、予備実験のサンプルでは全体の 25% がカラー成分を持っていた。そこで、画像の gray 値を式 (2) で求める。

$$gray = |R - G| + |G - B| + |B - R| \quad (2)$$

ここで、R は画像の赤成分、G は緑成分、B は青成分を表す。gray 値が 0 なら、その画像はグレーであることがわかる。

3.3 分類器作成

深層学習以外の機械学習の手法については、[11] を参考にして、Weka*4 を用いる。Weka は、インタフェースがわかりやすく多種の分類器を結合することやハイパーパラメータの変更が簡単に行える。また、混同行列をはじめ、各種評価指標が出力され、分類器の比較が簡単に行える。深層学習の手法については、[12] を参考にして、バックグラウンドが TensorFlow である Keras を用いる。

3.4 分類器の評価指標

分類器の性能を調べるためには混同行列 (表 1) がよく用いられるが、多数の分類器を比較したい場合には不向きである。そこで各分類器の性能を表す評価指数として、デー

*3 元の文書中では「正規化ガンマ関数」を表している。

*4 <https://www.cs.waikato.ac.nz/ml/weka/index.html>

表 1 混同行列

		分類器の予測	
		Positive	negative
実際の クラス	Positive	TP(True Positive)	FN(False Negative)
	Negative	FP(False Positive)	TN(True Negative)

タセットが不均衡データ*5であることを考慮して、マシューズ相関係数 (Matthews Correlation Coefficient: MCC) (式 (3)) を用いる。Chicco らは「高品質のスコアを取得するには、データセット全体の比率に関係なく、分類器は負のケースの大部分と正のケースの大部分の両方で正しい予測を行う必要がある。(略) 混同行列を使用するすべての研究者は、F 値や精度を使用する代わりに、MCC を介して二項分類予測を評価することを勧める [13]」と述べている。また、Yao らも [14] で、F 値と比較した MCC の利点について述べている。F 値 (式 (4)) も広く用いられているので、評価指標としてあげる。

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (3)$$

$$F\text{-measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

ここで TP, FP, FN, TN は表 1 中のものであり、

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN} \text{ である。}$$

3.5 提案手法の概要

意味のある数式がどのような特徴量を持つかを 4.2 で確定した後、分類器の性能を MCC を用いて評価する。実験 1 では、深層学習以外の機械学習の方法を用い、実験 2 では、深層学習の方法を用いて実験し、MCC を用いて結果を比較する。

4. 実験

4.1 データセット

先行研究 [15] のデータセットを使用する。ピシヨップの『パターン認識と機械学習』上下巻 [16][17] の索引から無作為に 100 キーワードを選び、それらのキーワードをクエリとして Web 検索を行い、上位 100 件の Web ページを取得した。その中で、Web 数式文書から数式画像を抽出してデータセットを作成した。キーワード 1 から 30 までのデータセットを D_v 、31 から 60 までを D_{tr} 、61 から 100 を D_{ts} とする。 D_v は検証用に、 D_{tr} は学習用に、 D_{ts} はテスト用に使用する。

4.2 「意味のある数式」条件決定

D_{tr} の正解画像 314 個を分析する。[15] では、クエリと

*5 正解は全体の 25%

表 2 特徴量決定

	A	B	C	D	E
全画像数 314	関係式 295	式数1 252 式数2 33 計 285	連結数1 241 連結数2 34 計 275	連結数3 6 連結数4 1 連結数5 2 連結数6 0 連結数7 1	行数1 260
					行数2 14
					計 274
					行数3 1
					式数3 5
					式数4 4
					式数5 0
					式数6 1
					関係式以外 19
					計

表 3 CNN 以外で使用するデータ内訳

	D_v	D_{tr}	D_{ts}
正解	2,410	1,868	4,336
不正解	2,562	6,076	12,689
計	4,972	7,944	17,025

表 4 CNN で使用するデータ内訳

	D_v	D_{tr}	D_{ts}
正解	2,400	1,900	4,462
不正解	2,400	1,900	18,678
計	4,800	3,800	23,140

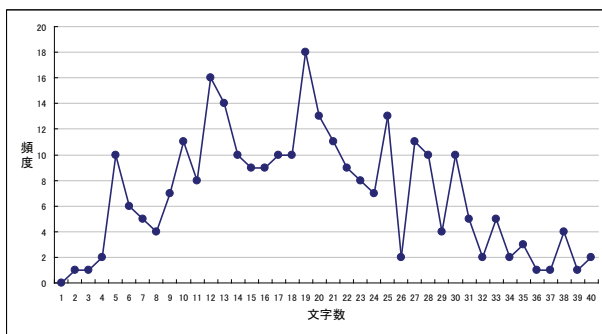


図 2 横文字数

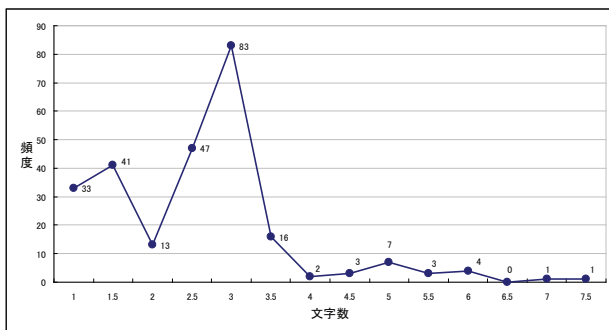


図 3 数式高さ文字数

して与えられたキーワードに関連する画像が正解画像である。パターン認識や機械学習に関連する数式画像は、意味のある数式である可能性が高いと考えられる。

「横文字数」, 「数式高さ文字数」, 「行数」, 「式数^{*6}」, 「連結数 (=, < などの数)」について調べ、特徴量の値を決定する。横文字数はフォントの大きさに関係なくカウントする。また例えば $\frac{x-1}{x^2+x-1}$ なら多いほうで 6 とする。数式縦文字数は指数は 0.5 としたが、最終的にはベースのフォントの大きさから目視で決める。項目が多いので、閾値の決定が容易なものから順次行った。表 2 で B 列の各値から、まず関係式とする。関係式とは、左辺と右辺が「=」,

*6 () 中は 1 つとする 式中の入れ子の式はノーカウント

「<」^{*7}などで結ばれているものである。次に、C 列から式数を 1 または 2 とし、D 列から連結数を 1 または 2 とし、E 列から行数を 1 または 2 とする。図 2 より、文字数 4 から 35 の間が頻度 2 以上なので、横文字数は 4 から 35 とした。図 3 より、文字数 6 以下が頻度 2 以上なので、高さは 6 以下とした。よって、意味のある数式は次の条件を満たすものとする。

- (1) 関係式 (2) 横文字数は 4 以上 35 以下
- (3) 数式高さ文字数は 6 以下 (4) 式数は 1 または 2
- (5) 連結数は 1 または 2 (6) 行数は 1 または 2

4.3 前処理

元データにはアイコンなどの不要な画像があるので、[15]の方法を用いてあらかじめ削除する。 D_{tr} の不正解画像名から単語を抜き出したものの内、正解画像名から単語を抜き出したものと一致するものを省いた残りを頻度順に並び替え、上位 129 個を不要語リストとした。 D_v に対して、画像名中に不要語リスト中の単語を持つものを省いた。

次に D_v と D_{tr} にはその作成方法から、同じ画像が違う ID で登録されているものを含んでいるので、特徴量が重複するものを削除する。CNN では直接画像を利用するので、すべての特徴量が同じものも最終的には目視で違いがあるかどうか確認した。

また、CNN で使用するには α チャンネルを削除する必要があり、その処理も行った。

4.4 正解数式の決定

4.2 の条件に従って、 $D_v \cdot D_{tr} \cdot D_{ts}$ の正解・不正解を手で調べる。表 3 は CNN 以外で使用する前処理済みのデータの内訳である。 D_{tr} はアンダーサンプリングで正解・不正解のバランスを取っている。表 4 は CNN 用のデータの内訳であり、 D_{tr} はアンダーサンプリングを行い、 D_{tr} と D_v はバッチ処理のため数を調整している。また、前処理の違いにより表 3・表 4 の数は一致していない。

*7 「等号」 299 個, 「不等号」 14 個, 「 \approx 」 4 個, 「 \simeq 」 4 個, 「 \equiv 」 2 個, 「:=」 2 個, 「 \sim 」 2 個, 「 \gg 」 1 個

表 5 CNN 以外の機械学習による分類器作成 : 各特微量上位 1 件

特微量	分類器	MCC	F 値
A 基本特微量	SVM	0.543	0.654
B 基本特微量 密度	SVM	0.577	0.677
C 基本特微量 横縦比	SVM	0.558	0.661
D 基本特微量 密度 横縦比	SVM	0.564	0.669
E 基本特微量 gray 値	MultilayerPerceptron	0.529	0.638
F 基本特微量 密度 gray 値	RandomSubSpace	0.538	0.649
G 基本特微量 横縦比 gray 値	MultilayerPerceptron	0.526	0.635
H 基本特微量 密度 横縦比 gray 値	RandomSubSpace	0.546	0.653
I 基本特微量+gray 値	SVM	0.545	0.658
J 基本特微量 密度+gray 値	SVM	0.570	0.677
K 基本特微量 横縦比+gray 値	SVM	0.533	0.623
L 基本特微量 密度 横縦比+gray 値	SVM	0.536	0.626

表 6 CNN 以外の機械学習による分類器作成 : 上位 3 件

特微量	分類器	MCC	F 値
B 基本特微量 密度	SVM	0.577	0.677
J 基本特微量 密度+gray 値	SVM	0.570	0.677
D 基本特微量 密度 横縦比	SVM	0.564	0.669

4.5 実験 1 深層学習以外の機械学習の手法

ここでは、画像の特微量を用いて意味のある数式とそれ以外を分類する分類器を作成し、それらの性能を MCC を用いて比較する。基本特微量を「ファイルサイズ」、「ファイル横幅ピクセル数」、「ファイル高さピクセル数」とし、さらに「密度 (式 (5))」、「横縦比 (式 (6))」、「gray 値 (式 (2))」を用いる。

$$\text{密度} = \frac{\text{ファイルサイズ}}{\text{横幅ピクセル} \times \text{高さピクセル}} \quad (5)$$

$$\text{横縦比} = \frac{\text{ファイル横幅ピクセル}}{\text{ファイル高さピクセル}} \quad (6)$$

特微量の条件を様々に変え、Weka で多種の分類器を作成する。表 5 に D_v を用いた結果を示す。表中「+gray 値」は、gray 値以外の特微量で作成した分類器の予測値が正解でかつ gray 値が 0 であるものを改めて予測値として計算したものである。いずれの分類器も、10 分割交差検証を行っており、ハイパーパラメータを探索している。多層パーセプトロン (Multilayer Perceptron) は中間層 (隠れ層) の数について、ランダム部分空間 (Random Subspace) [18] はアンサンブル学習であるので使用する分類器について、複数検討した。また、SVM は RBF カーネルを用いた非線形 SVM を使用し、 c と γ の値を探索した。

表 5 中、上位 3 件を表 6 に示す。MCC の値が一番良い B について D_{ts} を用いて評価実験を行った結果、MCC が **0.504**、F 値が **0.628** であった。なお、ここで使用した SVM は $c = 7000$ 、 $\gamma = 6$ である。

4.6 実験 2 深層学習の手法

ここでは、深層学習の手法の 1 つである CNN を用いて、意味のある数式とそれ以外の二値分類を行う。以前は、深層学習を行うには大量の画像データが必要とされ、この点がボトルネックとなっていたが、最近では転移学習用の学

表 7 CNN による二値分類 : 転移学習使用一覧

転移学習 モデル	データ 拡張	ファイン チューニング	MCC	F 値
A baseline	×	×	0.700	0.752
B baseline	○	×	0.590	0.649
C VGG16	○	○	0.719	0.765
D VGG16	×	×	0.760	0.765
E VGG16	×	○	0.771	0.814
F VGG16	部分○	○	0.712	0.758
G VGG19	○	○	0.668	0.720
H Xception	○	○	0.665	0.725
I InceptionV3	○	○	0.654	0.718
J NASNetMobile	○	○	0.675	0.730
K MobileNetV2	○	○	0.719	0.773
L MobileNet	○	○	0.735	0.782
M MobileNet	×	×	0.335	0.450
N MobileNet	×	○	0.743	0.788
O DenseNet121	○	○	0.734	0.784
P DenseNet121	×	×	0.371	0.489
Q DenseNet121	×	○	0.700	0.749
R DenseNet121	部分○	×	0.280	0.430
S DenseNet121	部分○	○	0.727	0.772
T DenseNet169	○	○	0.730	0.781
U DenseNet201	○	○	0.713	0.767
V ResNet50	○	○	0.693	0.747
W ResNet50_20epochs	○	○	0.711	0.761
X InceptionResNetV2	○	○	0.670	0.722

習済みモデルが広く公開され、一般に使用可能となっている。そこで、CNN でよく使われる「データ拡張」、「転移学習」、「ファインチューニング」の手法を用いて、意味のある数式の二値分類を行う。CNN による画像認識全般については [19] で、「データ拡張」、「転移学習」、「ファインチューニング」については、[20] で解説されている。

表 7 は、表 4 のデータを用いて、学習・検証・テストを行った最終のテストデータの結果である。最初にベースラインとして、Conv2D 層が 4 つ、MaxPooling2D 層が 4 つの A を構築する。活性化関数は sigmoid 関数、損失関数は binary_crossentropy である。B はさらにデータ拡張を行っている。データの数少なすぎることが原因で過学習が起こるのを避けるため、通常行われるデータの増し作業である。ここでは、画像のランダムな回転、平行移動、等積変形、ズーム、半分の水平方向に反転、を行った。例を図 4 に示す。(a) は D_{tr} の 1 番目の画像であり、(b) はデータ拡張によって (a) から作られた画像である。表 7 の C から W は、Keras のライブラリにある学習済みのネットワークを使用した CNN である。それぞれの名前の後の数字はおおよそ層の深さを表す、基本パターンは、データ拡張、転移学習用のモデルの使用、ファインチューニングのすべてを行い、出力した予測値に基づいて二値分類したテストデータの統計量を計算するものである。エポック数 (処理の繰り返し数) は 30 とした。学習と検証では、途中経過を調べるために、精度グラフと損失グラフを出力している。W については、精度グラフで 20 エポックあたりから急激に

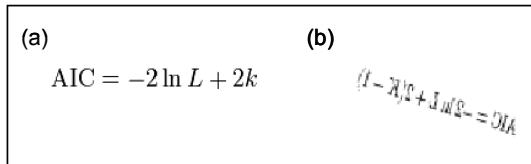


図 4 データ拡張

表 8 CNN による二値分類：上位 3 件

転移学習 モデル	データ 拡張	ファイン チューニング	MCC	F 値
E VGG16	×	○	0.771	0.814
D VGG16	×	×	0.760	0.765
N MobileNet	×	○	0.743	0.788

精度が落ちていることがわかり、処理を 20 エポックで止めたものである。

基本パターンの中で、MCC の値が良く、処理速度もそれほどかからないものについては、条件を変えて実験を繰り返した。まず、データ拡張については、数式が反転したり傾いたりしていることは稀なので、平行移動のみに限ったものが、表 7 中のデータ拡張列の「部分○」である。しかし、この操作はあまり有効ではなかった。ファインチューニングに関しては、Keras の作者である Chollet が「ファインチューニングは、特徴抽出に使用される凍結された畳み込みベースの出力側の層をいくつか解凍し、モデルの新しく追加された部分と解凍した層の両方で訓練を行う、という仕組みになっている。この手法が『ファインチューニング』と呼ばれるのは、現在取り組んでいる問題への関連性を高めるために、再利用するモデルのより抽象的な表現を微調整するからだ [12]」と述べている。表 7 を見ると、データ拡張を行わないことによって結果が悪くなった場合でも、ファインチューニングによって結果が向上することがわかった。

表 7 中の MCC に関して上位 3 件を、表 8 に示す。4.5 の最終結果 (MCC が **0.504**, F 値が **0.628**) と表 8 の MCC 値を比べると、データ拡張をせずに VGG16 による転移学習とファインチューニングを行った CNN が一番良いことがわかった。処理速度に関しても、SVM は、ハイパーパラメータの値が大きいため時間がかかるのに対し、CNN のほうは層数が少なくそれほどかからない。

5. 応用

4 章の結果を、先行研究に応用してみる。先行研究 [15] は、数学用語をクエリとして Web 検索を行い、その用語と関連する数式を上位 10 件提示するというものである。

各画像 i_k に与える点数は以下の式 (7) で与えられる。

$$score(i_k) = x_{line} + x_{key} + x_{svm} + x_{bo} \quad (7)$$

ここで、 i_k が独立行にあれば x_{line} が 1、それ以外は 0、キーワードをウィンドウサイズ 100 文字内で持てば x_{key} が 1、

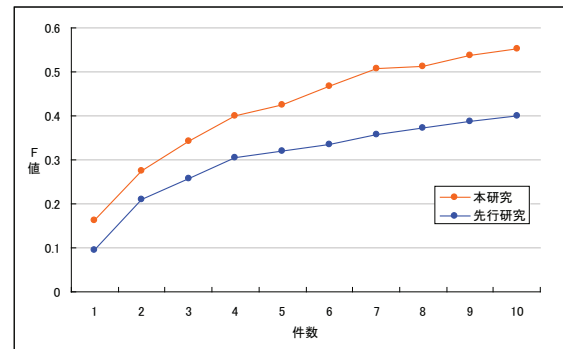


図 5 分類器を取り換えた F 値の比較

それ以外は 0、分類器による正解数式らしさの予測値が 1 なら x_{svm} が 1、それ以外は 0 とする。ここまでで、満点は 3 点となる。もし、 i_k が、満点を持ちかつ同一の Web 文書内で最初に出現したなら、ボーナス点として x_{bo} を 1 とし、それ以外を 0 とする。

分類器には Chang による Libsvm[21] を用いた。RBF カーネルによる非線形の SVM である。この時の正解は、キーワードに関連している数式であり、本研究とは異なる。また、不正解が非常に多い不均衡データ*8であったため、学習データの不正解をアンダーサンプリングしている。

評価指標として F 値 (式 (4))・平均逆順位 (Mean Reciprocal Rank: MRR) (式 ((8)))・平均精度の平均 (Mean Average Precision: MAP) を用いる。各キーワードについての上位 10 件の平均精度 (Average Precision: AP) (式 (9)) で、 $md = 10$ として計算し、さらにそのマクロ平均を計算することで求める。F 値 (式 (4)) については、 $Precision@n = \frac{r}{n}$, $Recall@n = \frac{r}{c}$ である。

$$MRR = \frac{1}{n} \sum_{k=1}^n \frac{1}{r'_k} \quad (8)$$

$$AP@md = \frac{1}{\min(md, c)} \sum_s I(s) Prec(s) \quad (9)$$

ここで、 r は上位 n 件の正解画像数、 c は全正解画像数、 r'_k は k 番目のキーワードにおける最上位にある正解の順位、 md は測定長、 $I(s)$ は第 s 位の画像が正解か否かを表すフラグ、 $Prec(s)$ は第 s 位の精度である。

式 (7) 中の x_{svm} を MCC の値が一番良かった、データ拡張せずに VGG16 による転移学習とファインチューニングを行った CNN に取り換えて、再度実験を行った。結果は、先行研究の MAP 値が **0.42**, MRR 値が **0.84** に対し、本研究では、MAP 値が **0.45**, MRR 値が **0.86** となった。図 5 に F 値の比較を示す。グラフより F 値も本研究が先行研究を上回っていたことがわかる。

6. おわりに

本研究は、Web 数式文書中にある多数の画像から、式

*8 テストデータ中、正解は全体の 7%弱

(1) のような「コンパクトに表現された単独で意味のある数式」を抽出することを目指すものであり、意味のある数式とその他の画像を教師あり機械学習の方法で二値に分類することによって実現した。

最近では、Web 文書中の数式の表現方法として MathML 等が使われることも多いが、まだまだ多くの数式画像が存在する。深層学習では大量のデータが必要とされ、データ収集のコストがボトルネックとなっていたが、少数のデータであっても転移学習やファインチューニングを行うことによって良い結果を得られることを示せた。VGG16 は ImageNet という 1400 万枚以上のカラー写真で構成されたデータセットで学習された 16 層からなる CNN モデルである。それらのカラー写真は、風景・動物・人・静物・乗り物など様々である。一方で、数式画像はそれらとはかけ離れているのに、従来の機械学習の方法を上回る結果を出すことができた。

また、意味のある数式を定義することにより、先行研究で今後の課題とした「SVM の限界を克服するために Deep Learning などの機械学習の手法を用いて数式画像を判別し出力結果の向上を目指す」ことを実現することができた。

さらにより結果をめざすために、従来の機械学習の方法と CNN を組み合わせるなどのアンサンブル学習について考察を進めることを今後の課題としたい。

参考文献

- [1] Aizawa, A., Kohlhase, M. and Ounis, I.: 数式検索タスク NTCIR-11 Math-2, 第 5 回人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会 (SIG-AM), 05-11, pp. 58-60 (2013).
- [2] Zanibbi, R. and Blostein, D.: Recognition and Retrieval of Mathematical Expressions, *International Journal on Document Analysis and Recognition*, Vol. 15(4), pp. 331-357 (2012).
- [3] Aizawa, A. and Kohlhase, M.: Mathematical Information Retrieval, in *Evaluating Information Retrieval and Access Tasks*, Springer, Singapore, 169-185 (2021).
- [4] Hussain, S. and Khoja, S.: Retrieval of Mathematical Information with Syntactic and Semantic Structure over Web, *Journal of Information Science & Engineering*, Vol. 36 (1), pp. 75-89 (2020).
- [5] Zhong, W. and Zanibbi, R.: Structural Similarity Search for Formulas using Leaf-Root Paths in Operator Subtrees, *Advances in Information Retrieval*, Springer, pp. 116-129 (2019).
- [6] Greiner-Petter, A., Schubotz, M., Müller, F., Breitingner, C., Cohl, H., Aizawa, A. and Gipp, B.: Discovering Mathematical Objects of Interest-A Study of Mathematical Notations, *Proceedings of The Web Conference 2020 (WWW'20)*, pp. 1445-1456 (2020).
- [7] Ohashi, S., Kristianto, G. Y., Topic, G. and Aizawa, A.: Efficient Algorithm for Math Formula Semantic Search. *IEICE Transactions on Information and Systems*, Vol. 99, Issue 4, pp. 979-988 (2016).
- [8] Suzuki, T. and Fujii, A.: A Structure-based Method for Mathematical Document Classification, *IPSSJ SIG Technical Report*, Vol. 2017-IFAT-125, No.11, pp. 1-8, (2017).
- [9] Shirmenbaatar, M., 古賀 久志, 渡辺 俊典: 数式画像をクエリとする類似数式検索システム, 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM2012), E6-2 (2012).
- [10] Rani, S. J. and Kumari, V. V.: An Effective Mechanism of Feature Based Retrieval of Mathematical Expression from Documents, *International Journal of Applied Engineering Research*, Vol. 11, No. 5, pp. 3462-3468 (2016).
- [11] 荒木 雅弘: 第 1 版 フリーソフトではじめる機械学習入門, 森北出版株式会社 (2014).
- [12] François Chollet, 株式会社クイープ訳, 巢籠 悠輔監訳: 初版 Python と Keras によるディープラーニング, 株式会社 マイナビ出版 (2018).
- [13] Chicco, D., Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6 (2020).
- [14] Yao, J. and Shepperd, M.: Assessing Software Defection Prediction Performance: Why Using the Matthews Correlation Matters, *Proceedings of the Evaluation and Assessment in Software Engineering (EASE '20)*, pp. 120-129 (2020).
- [15] 山田 奉子, 上田 洋, 村上 晴美, 岡 育生: 数学用語をクエリとする Web 上の数式画像検索, *人工知能学会論文誌* 33 (4) P. A-H911.1-13, (2018).
- [16] ビショップ, C. M., シュプリングー・ジャパン株式会社編, 元田 浩, 栗田 多喜夫, 樋口 知之, 松本 裕治, 村田 昇監訳: パターン認識と機械学習 上, 丸善出版株式会社 (2012).
- [17] ビショップ, C. M., シュプリングー・ジャパン株式会社編, 元田 浩, 栗田 多喜夫, 樋口 知之, 松本 裕治, 村田 昇監訳: パターン認識と機械学習 下, 丸善出版株式会社 (2012).
- [18] Ho, T. K.: The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 20, No.8, pp. 832-844 (1998).
- [19] 矢田真城, 魚住龍史: 畳み込みニューラルネットワークによる画像認識, *情報処理学会研究報告*, Vol. 2020-BIO-62, No.8, pp. 1-6, (2020).
- [20] 中山英樹: 深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習, *電子情報通信学会技術研究報告*, Vol. 115, No. 146, pp. 55-59 (2015).
- [21] Chang, C.-C. and Lin, C.-J.: LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology*, 2, 27:1-27:27 (2011).