

数式検索システムを応用した学習項目 ならびに公式抽出機能の実装

粥川 佳哉^{*1}, 宮崎 佳典^{*2}

^{*1} 静岡大学情報学部

^{*2} 静岡大学大学院情報学領域

Extraction of Mathematical Concepts and Formulas as Application of Math IR System

Yoshiya Kayukawa ^{*1}, Yoshinori Miyazaki ^{*2}

^{*1} Faculty of Informatics, Shizuoka University

^{*2} Graduate School of Informatics, Shizuoka University

数式には、分数や行列などといった、記号が水平に並ばない構造が存在する。このような数式を計算機上で表現するデータ形式として MathML や TeX などが挙げられる。これらのデータ形式の普及に伴い数式のデータが蓄積され、数式に対する検索技術の必要性が高まっている。これに対し、我々は数式を計算機上で表現するデータ形式の一つである MathML Presentation Markup に着目し、正規表現を用いた数式検索システムを開発してきた。同システムの応用として、学習支援の側面から学習項目抽出機能の実装にも着手し、学習者にとって初見の数式に対し数式理解のヒントや指針を与える機能をシステムに追加する。さらに、事前に登録した公式情報から公式名のみで数式を検索する機能の実装についても触れる。

キーワード: 数式検索, 学習支援, 学習項目, 公式, MathML

1. 導入

数式には、分数や行列などといった、記号が水平に並ばない構造が存在する。こうした構造を含めた数式を計算機上で表現するデータ形式として MathML(1) や TeX などが挙げられる。また、これらのデータ形式の普及に伴い数式データが蓄積されたことで、数式に対する検索技術の必要性が高まっている。ただ、構造が複雑なだけに検索システムそのものも複雑になり、さらにクエリを入力する側も数式入力是一般に困難になる。先行研究として、宮崎らは MathML のタグセットの 1 つである MathML Presentaion Markup に着目し、正規表現を用いた数式検索システムを開発してきた(2)。また同時に (2)ではシステムの応用として、数式に含まれる学習項目を抽出する機能も実装してき

た。学習者が初めて見る数式に直面した場合、この機能を用いることで何を学習すればその数式を理解できるのかを知ることができる。これは学習支援システムとしての利用の糸口になると考えられる。一方で、(2)における学習項目抽出機能は、実装が行列に関連した学習項目を中心に実装されており、抽出可能な学習項目は非常に限られていた。よって本研究では、教育支援の観点から、先行研究で開発された学習項目抽出機能付き数式検索システムに対し、学習項目の抽出範囲の拡大を目指す。また、学習項目抽出機能の適用例として新たに、公式検索機能の実装を行う。これは、事前に登録した定型の公式情報から、公式名を与えることでその公式を含む(あるいは公式そのものの)数式、あるいはその公式を用いて数式変形している個所を検索する機能である。または、ある数式変形がその公式

を適用したものである場合、それを学習者に提示することができる。一般に、学習テキスト中における式変形の根拠や元になる概念は必ずしもテキスト内に明示されるわけではなく、本機能はテキストを理解するための一助にもなり得ると考えられる。

2. 数式検索システム

本研究で扱うシステムは、現在 Firefox のバージョン 49.0.2 上で動作し、MathML Presentation Markup で記述された数式を対象に検索を行うものである。MathML Presentation Markup はタグを用いて記号の配置を指定できるタグセットである。例として、mi 要素は識別子、mo 要素は演算子、mn 要素は数を表す。また、MathML データは入れ子構造を持っており、入れ子の外側を親、内側を子とすることで木構造として表現することができる。例として、 $(\frac{x}{2})^2$ の MathML Presentation Markup のデータと、そのデータを木構造として表現したものを表 1 に示す。なお、左側が MathML Presentation Markup のデータ、右側がそのデータを木構造で表現したものである。

表 1 $(\frac{x}{2})^2$ の MathML データと木構造による表現

<code><math></code>	math
<code><msup></code>	└─msup
<code><mfenced></code>	├─mfenced
<code><mfrac></code>	│ └─mfrac
<code><mi>x</mi></code>	│ ├─mi/x
<code><mn>2</mn></code>	│ └─mn/2
<code></mfrac></code>	└─mn/2
<code></mfenced></code>	
<code><mn>2</mn></code>	
<code></msup></code>	
<code></math></code>	

なお、mfenced 要素はその子を括弧記号で囲むための要素、mfrac 要素は第一の子を分子、第二の子を分母として記号を配置する要素である。数式検索システムで MathML Presentation Markup を扱うためには、データのゆらぎを解消する必要がある。データのゆら

ぎとは、MathML Presentation Markup において同一の数式を表すデータが一意に定まらないことを指す。例として、mi 要素に関するものが挙げられる。mi 要素は複数の文字を子に持つことが可能であるため、例えば正弦を表す sin を MathML Presentation Markup を用いて表したい場合には、

`<mi>sin</mi>`

や

`<mi>s</mi><mi>i</mi><mi>n</mi>`

といった複数種類の記述が考えられ、この性質はデータのゆらぎの原因となる。こういったゆらぎを解消するため、著者らが開発している数式検索システムでは、正規化処理を行うことで数式の表現を統一している。また、クエリ内に正規表現を許すことで曖昧検索機能を実現している。検索の結果、一致した部分をハイライト表示または置換を行うことが可能である。

数式検索システムの入力インターフェースを図 1 に示す：

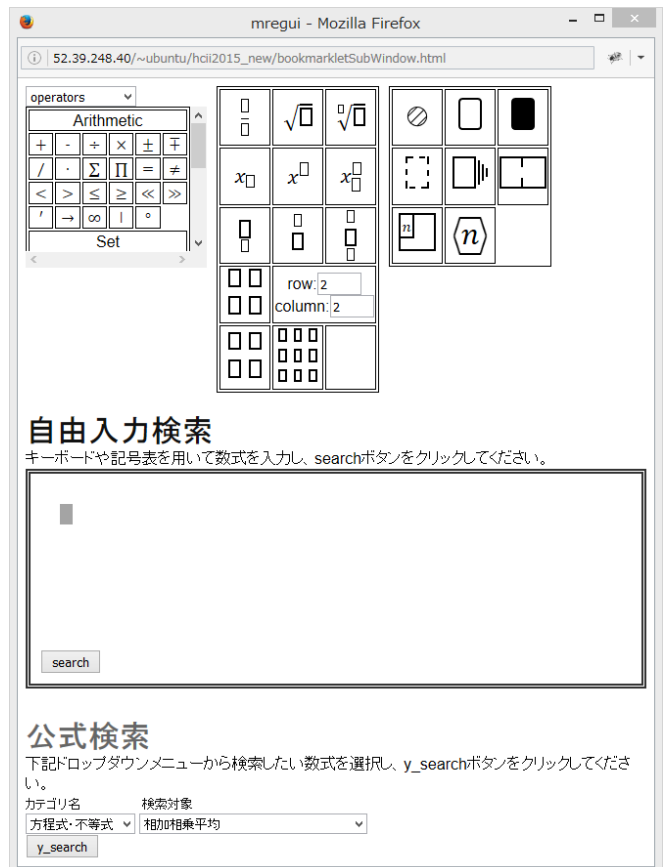


図 1 数式検索システムの入力インターフェース

インターフェース下部「自由入力検索」内に、クエリ用数式を入力する。上部には数学記号、数学特有の構

造あるいは正規表現入力パレットが配置されている。左上の記号群は数学で用いられる記号がまとめられており、クリックすればその記号が入力される。中央のアイコンは分数や上下付き文字、行列といった数式の構造を記述するためのものである。右上の8つのアイコンは正規表現に関わるもので、表2に各アイコンの詳細を示す。

表2 正規表現一覧

	任意の1文字		内部入力文字のいずれか		内部入力文字以外
	内部の文字列と最大1回マッチ		入力文字列の1回以上の繰り返し		仕切りごとの入力文字列のいずれか
	後方参照用のラベル(n)		n(番)を後方参照		

ユーザーは図1上部のパレット内の要素を自由に組み合わせることで、複雑な曖昧検索を実現することが可能となっている。クエリを作成し終えた後に search ボタンをクリックすることで数式が検索されることとなる。

一例として、“相加・相乗平均の式”を検索したい場合を考える。

図2 自由入力検索の入力例（相加・相乗平均）

図2が“相加・相乗平均の式”である。表2より、が任意の1文字、が内部入力文字列の1回以上の繰り返しのため、は任意の文字列（数式）を表す。これに後方参照用に1番のラベルを割り振っている。同様に、これとは別に任意の文字列（数式）に2番のラベルを割り振っている。これが図2の左辺の分子で用いられている正規表現である。右辺は①で先ほどの1番を後方参照している。2番についても同様である。よってこの部分には左辺で用いられた2つの任意の数

式を表している。次にについて説明する。まず外枠のは“内部の入力文字列と最大1回マッチ”，つまり内部の文字列にマッチしてもしなくてもよい（する時は一度のみ）。その内側にあるのがであり、これは“内部入力文字のいずれか”を表す。その中に“×”と“・”があるため、“×”か“・”のいずれかと最大1回マッチすることを表す。以上をまとめ、は“×”か“・”が入るか、あるいは何も入らないことを示す。

2数 a, b の乗算を表現する際には“ ab ”や“ $a \times b$ ”，“ $a \cdot b$ ”が用いられることが多いため、このような表記を用いている。つまり、図2は“ある数式（1番）とある数式（2番）を加算して2で割ったもの \geq 平方根（1番の数式と2番の数式の乗算）”となり、“相加・相乗平均の式”を表す。これは例えば、

$$\frac{a+b}{2} \geq \sqrt{a \cdot b}$$

や

$$\frac{x_1+x_2}{2} \geq \sqrt{x_1x_2}$$

や

$$\frac{1}{x^2+x+1} + (x^2+x+1) \geq \sqrt{\frac{1}{x^2+x+1} \times (x^2+x+1)}$$

などがマッチする。一方で、相加・相乗平均の式に式変形を施した、

$$a+b \geq 2\sqrt{a \cdot b} \quad \text{や} \quad \frac{a+b}{2} - \sqrt{a \cdot b} \geq 0$$

は現時点ではマッチしない。

3. 公式検索

上記で説明した数式検索システムは、表2の正規表現を組み合わせることで、様々なパターンの数式を一般に検索させることができ、学習に有効と考えられる。これを応用したものが公式検索である。パターン化された公式を、数式検索システムで実装されている正規表現を用いて表現し、そのクエリをあらかじめ保存しておくことで、より容易にその公式の形を含んだ数式をユーザーに提示することができる。公式検索は、図1の下部にある「公式検索」から実行する。

以下に、現時点で登録されている公式一覧と、その公式のクエリ（表3）を示す。

登録されている公式

- 三角関数（正弦，余弦，正接）に関する加法定理，倍角公式，半角公式
- 三角関数（正弦，余弦）に関する三倍角公式
- 三角関数の積和公式，和積公式
- log の底の変換公式
- オイラーの公式
- コーシー・シュワルツの不等式（2次，3次）
- ド・モアブルの定理
- ピタゴラスの定理
- ヘロンの公式
- 相加・相乗平均の式
- 多重根号を外す公式

表3 登録されている公式に対応するクエリの例

加法定理（正弦）
$\sin(\overset{1}{\square}\overset{\ominus}{\square} + \overset{2}{\square}\overset{\ominus}{\square}) = \sin(\overset{1}{\square})\cos(\overset{2}{\square}) + \cos(\overset{1}{\square})\sin(\overset{2}{\square})$
三倍角公式（余弦）
$\cos 3\overset{1}{\square}\overset{\ominus}{\square} = 4\cos^3(\overset{1}{\square}) - 3\cos(\overset{1}{\square})$
積和公式（余弦×余弦）
$\cos\overset{1}{\square}\overset{\ominus}{\square} \times \cos\overset{2}{\square}\overset{\ominus}{\square} = \frac{1}{2} \times \{\cos(\overset{1}{\square} + \overset{2}{\square}) + \cos(\overset{1}{\square} - \overset{2}{\square})\}$
log 底の変換公式
$\log_{\overset{1}{\square}\overset{\ominus}{\square}} \overset{2}{\square}\overset{\ominus}{\square} = \frac{\log_{\overset{3}{\square}\overset{\ominus}{\square}} \overset{2}{\square}}{\log_{\overset{3}{\square}} \overset{1}{\square}}$
オイラーの公式
$e^{i\overset{1}{\square}\overset{\ominus}{\square}} = \cos(\overset{1}{\square}) + i\sin(\overset{1}{\square})$
コーシー・シュワルツの不等式
$(\overset{1}{\square}\overset{\ominus}{\square}^2 + \overset{2}{\square}\overset{\ominus}{\square}^2)(\overset{3}{\square}\overset{\ominus}{\square}^2 + \overset{4}{\square}\overset{\ominus}{\square}^2) \geq (\overset{1}{\square}\overset{3}{\square} + \overset{2}{\square}\overset{4}{\square})^2$
ド・モアブルの定理

$(\cos\overset{1}{\square}\overset{\ominus}{\square} + i\sin(\overset{1}{\square}))^{\overset{2}{\square}\overset{\ominus}{\square}} = \cos(\overset{2}{\square}\overset{1}{\square}) + i\sin(\overset{2}{\square}\overset{1}{\square})$
ピタゴラスの定理
$\overset{\ominus}{\square}\overset{\ominus}{\square}^2 = \overset{\ominus}{\square}\overset{\ominus}{\square}^2 + \overset{\ominus}{\square}\overset{\ominus}{\square}^2$
ヘロンの公式
$\overset{\ominus}{\square}\overset{\ominus}{\square} = \sqrt{\overset{1}{\square}\overset{\ominus}{\square}(\overset{1}{\square} - \overset{\ominus}{\square})(\overset{1}{\square} - \overset{\ominus}{\square})(\overset{1}{\square} - \overset{\ominus}{\square})}$
相加・相乗平均の式
$\frac{\overset{1}{\square}\overset{\ominus}{\square} + \overset{2}{\square}\overset{\ominus}{\square}}{2} \geq \sqrt{\overset{1}{\square} \times \overset{2}{\square}}$
多重根号を外す公式
$\sqrt{\overset{1}{\square}\overset{\ominus}{\square} + \overset{2}{\square}\overset{\ominus}{\square} + 2\sqrt{\overset{1}{\square} \times \overset{2}{\square}}} = \sqrt{\overset{1}{\square}} + \sqrt{\overset{2}{\square}}$

表3中の「多重根号を外す公式」を実行した時の出力結果を図3に示す。網掛けになっている数式がマッチした数式である。上から3番目の数式のみマッチしていないが、これは正しく後方参照のアルゴリズムが機能していることを示している。

$$\sqrt{a+b+2\sqrt{ab}} = \sqrt{a} + \sqrt{b}$$

$$\sqrt{x+y+2\sqrt{xy}} = \sqrt{x} + \sqrt{y}$$

$$\sqrt{a+b+2\sqrt{ab}} = \sqrt{a} + \sqrt{y}$$

$$\sqrt{\frac{a}{c} + b + 2\sqrt{\frac{a}{c}b}} = \sqrt{\frac{a}{c}} + \sqrt{b}$$

図3 公式検索の出力結果例（多重根号を外す）

4. 学習項目の抽出

自分が入力した数式に、事前に登録しておいた式の特徴がマッチする時、その特徴（項目名）を出力することが可能である。これをさらにフレキシブルに対応させれば、学習者にとって初見の定理・公式や数学の概念が含まれている数式に対して、それらを学習項目として表示することが可能となる。学習項目を抽出する際は、数式検索システムで正規化処理を行った数式を対象とする。学習項目に対応する正規化された数式のパターンの集合をあらかじめ保存しておき、それらにマッチするかを検索システムで逐次実行すればよいことになる。(2)では行列関連の学習項目を中心に実装していたが、新たに学習項目を追加実装した。追加した学習項目と抽出する際に着目した特徴を表4に示す。

表4 新たに実装を行っている学習項目

学習項目	特徴
微分	ライプニッツの記法, ラグランジュの記法
積分	積分記号 (∫)
極限	極限記号 (lim)
方程式	等号 (=)
不等式	不等号 (<, >, ≤, ≥)
二次式	次数 (2)
絶対値	絶対値記号 ()
平方根・累乗根	記号 (√, ⁿ√)

“方程式”の抽出の特徴として用いている等号“=”は、総和や総乗を表す記号“Σ”や“Π”の下にも記述される場合がある。そのため、現在は総和・総乗を表す記号の下に記述された等号からは方程式を抽出しないようにしている。また、複数の学習項目によって新たな学習項目が生成される場合、各々の学習項目に合致すればよいため、既存の学習項目を含んだ階層構造を構築し、“行列方程式”（“行列”と“方程式”の学習項目に相当する数式パターン）や“微分方程式”（“微分”と“方程式”の学習項目に相当する数式パターン），“二次関数”（“二次式”と“関数”の学習項目に相当する数式パターン）などの実装を行っている。図4に、入力した数式クエリと対応する学習項目を出力した例

を示す。

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} + \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 4 \\ 2 & 6 \end{pmatrix}$$

抽出した学習項目:行列 正方行列 対称行列 行列方程式

図4 学習項目の抽出機能の出力例

5. 現行システムの課題

現行の数式検索システムに含まれている各機能の説明を2-4節にわたって述べたが、未解決となっている課題もいくつか存在する。本節ではその各課題について言及する。

まず本システムの根幹となっている数式検索処理部に関する問題点について触れる。システム上で表現できる構造の中に行列があるが、現状では行列のサイズに関する曖昧検索を行うことができない。例えば、大きさを指定して“3 次のスカラー行列”を検索することは、既存の正規表現を用いることで可能だが、サイズを問わずすべてのスカラー行列を検索したい場合には行列のサイズに関する曖昧検索が求められる。このような検索が可能となれば行列に関する検索の利便性は大きく向上すると考えられるが、スカラー行列や対称行列といった条件（すなわち、既存の正規表現を用いた曖昧検索）を満たした上で行列のサイズを問わない検索を行うことは難しく、現在は実装を行うに至っていない。

次に学習項目抽出機能に関する課題について説明する。学習項目は公式とは異なり、必ずしも定型が存在するわけではないため、指定した条件を満たすかどうかの判断が難しく、簡単に解決できない学習項目も一定数存在する。数式“ $\frac{dy}{dx} - 2y = 0$ ”は微分方程式であるが、同じく、微分作用素と等号を併せ持っている数式“ $\sum_{k=1}^n f_k'(x)$ ”は微分方程式ではない。また“ $y = ax^2 + bx + c$ ”は二次関数であっても、“ $y = a^2 + bx$ ”が二次関数か否かの判断を正しく行うことができない。また、数学において、一種類の表現が色々な場面で使われることは珍しくないため、その場合の学習項目の抽出はとても困難だと言える。例として、記号“’”

が挙げられる。現在は学習項目の一つである微分（ラグランジュの記法）を抽出する特徴として取り扱っているが、この記号は座標を表現する際に“点 O に対する点 O' ”など、微分以外でも用いられる場合がある。そのため現在は、厳密には記号“'”から学習項目を正確に抽出しているとは言えない。

次に公式検索に関する課題について説明する。公式検索は、定型の公式を検索システムにおける正規表現等を用いて表した“公式情報”を保存し使うことで検索を実行しているが、保存している公式情報と形が完全に一致している物しかマッチせず、現時点では、少しでも変形された公式はマッチしない（例として、現行のシステムで相加・相乗平均の式を公式検索した場合を挙げている）。

6. まとめと今後の展望

学習項目抽出機能の拡張として、行列に関連したもの以外の学習項目の実装を行った。これにより、学習支援システムとしての側面が強化されたと言える。また、抽出機能の適用例として新たに公式検索機能を実装した。定型パターンの公式に対する検索を実現し、ある数式が特定の公式を適用していることを学習者に提示することができるようになった。

本システムの今後の課題として、抽出可能な学習項目ならびに実行できる公式検索のさらなる拡張と、公式検索機能の実行時に許容する式変形のバリエーションに対する検討ならびに実装がある。学習項目抽出機能は現状では対応可能な学習項目の生成例が非常に少ないため、さらなる学習項目の拡張が必要だと考えている。公式検索機能も同様に対応できる公式の範囲を拡大することが最優先事項だと言える。これについては、実際に教育現場で用いられている教科書を参考にし、より実用的な抽出、公式検索を行えるよう改良を加えていきたいと考えている。公式検索はあらかじめシステムが保存している数式のパターンしか検索できず、両辺を交換したものや両辺を 2 倍したものなど、公式に式変形を行った形のものにはマッチしない。式変形を許容させる場合、そのバリエーションにどこまで対応させるか検討する必要がある。実装にあたってはあらかじめ保存しておく公式情報のさらなる拡充が求

められる。

参考文献

- (1) MathML, <https://www.w3.org/Math/>
- (2) Y. Miyazaki, K. Shinshi, Searching Mathematical Expressions with Regular Expressions Tool and Its Application to Extract Mathematical Concepts, Proceedings of Society for Information Technology & Teacher Education 2017 (SITE 2017), pp. 1974-1978 (2017).