

## MathML 記述の数式に対する学習項目の自動分類手法の実装

## Implementation of an Automatic Classification Method for Learning Items for Mathematical Formulas Written in MathML

BUI THI HA LAN<sup>\*1</sup>, 宮崎 佳典<sup>\*2</sup>Bu Thi Ha LAN<sup>\*1</sup>, Yoshinori MIYAZAKI<sup>\*2</sup><sup>\*1</sup>静岡大学 情報学部<sup>\*1</sup>Faculty of Informatics, Shizuoka University<sup>\*2</sup>静岡大学大学院 情報学領域<sup>\*2</sup>College of Informatics, Shizuoka University

Email: bui.thi.ha.lan.19@shizuoka.ac.jp

あらまし：当研究室では、MathML で記述された数式を対象に数式検索エンジンを開発している。本研究では、その応用機能の1つである学習項目抽出機能を用い、クエリとして入力された数式に対して、学習項目の自動分類ならびにグラフ描画の実装を行った。また、Wolfram Alpha API を適用することで数式の正規化処理の実装を行い、項の並び順や冗長な表記に起因する表記揺れに対処した。これにより、付随する項目抽出やグラフ描画の誤判定の軽減にも貢献した。

キーワード：数式検索システム、学習項目抽出機能、Wolfram Alpha API、グラフ描画、数式の正規化

## 1. 研究背景と目的

近年、数式用マークアップ言語 MathML などの標準化により、Web 上での数式表記の普及に加え、その応用として、数式検索や学習支援への道が拓けてきている。本研究では、MathML Presentation Markup [1]を対象とした正規表現ベースの数式検索のシステム [2]と、その学習支援拡張を継続的に展開してきた。

しかし、図形分野の数式は表記揺れや構文的多様性が大きく、既存機能では十分に対応できていなかった。また、式と図形対応の視覚支援も十分ではなかった。

本研究の目的は二つである。第一に、[1]の1応用例である学習項目抽出機能 [3]を図形分野へ拡張し、入力した数式クエリに対して対応グラフを適切に表示することである。第二に、入力数式の表記揺れに対応するため、Wolfram Alpha API [4] (以下 WA API) による代数的正規化を導入し、数式の図形カテゴリ分類に向けた安定性を高めることである。

## 2. 図形学習項目抽出機能

提案手法は、(A) 図形学習項目抽出機能(B) 数式正規化 (WA API 連携) および可視化から構成される (A)では、入力 MathML の構造を手がかりに、あらかじめ用意した「図形学習項目ごとの判定パターン一覧」に基づき分類する。この時、表記揺れを吸収できるよう、MathML トークンを前提とした共通パターン部を用いて正規表現を組み立てる。

MathML 式の構造的特徴を扱うためには、式を DOM 木 (expression tree) として表現し、構造的な類似性やパターンを捉えることが有効と報告されている[5]。また、そこで、本研究では、網羅性を重視し

て正規表現によるパターンマッチング[2]をまず行い、その結果に対して必要な型に限って DOM 解析による追加検証を行う。具体的には、分母が完全平方かどうかなどの与式の条件について、正規表現のみでは階層構造や指数表現を正確に判定できない場合に、DOM 構造を用いた形式的検証を行い、条件を満たす場合のみ採用することとする。

## 3. 数式正規化処理 (WA API 連携)

横井らは、数式検索では表記や木構造の一致だけでは不十分で、文脈や等価変形を考慮した意味的同値判定が重要であると指摘している[6]。

一方、本研究の図形学習項目の抽出は、主に「MathML の文字列パターン」と「構造」によって判定するが、項順差、指数冗長、展開/因数分解差、左右入れ替え、同類項整理・約分などがあると、パターン一致が困難になる。これは MathML Presentation Markup から MathML Content Markup への安定的な変換が依然として確立されていないという既存研究の指摘 [7] とも符合する。

そこで本研究では、表現依存の判定手法の限界を補完する目的で、(A)で処理しきれない場合に限り (送信の抑制=コスト・遅延対策) WA API を呼び出し、数学的に同値な式を標準形へ正規化する。主な呼び出し条件は (i) 同類項の重複・相殺や部分計算など「短くできる」可能性、(ii) 抽出が「なし/不明方程式」、(iii) 指数冗長パターン検出時である。

## 3.1. WA API で行う正規化

本研究では、方程式の形と変数に応じて、WA API への正規化クエリを切り替える。

(1) 「式 = y」は Collect [左辺, v] を利用し、(2) 「y = 式」は Collect [右辺, v] を用いて次数降順に整

理し、「 $y=式$ 」に統一する。(3)「 $式 = 式$ 」(一般型)  
 $y$  が非線形かつ暗黙に曲線の場合、図形表現を保つ  
 ため、式をそのまま WA API へ送る。一方、 $y$  が一次  
 である場合、(3) は  $Reduce[式, y]$  とすることで、「 $y$   
 $= 式$ 」への変換を許容する。

追加整理(文字係数を含む多項式):  $Collect$  と  
 $Reduce$  で次数降順に整列できない場合、  
 $MonomialList[Expand[式], x]$  のように整列して、「 $y =$   
 (和)」に戻すように工夫する。

#### 4. 図形可視化

図形学習分野においては、数式とそれに対応する  
 グラフ描画を同時に提示することで、その数式の理  
 解支援につながるが多い。

本研究では、検索結果でハイライトされた数式の  
 みを抽出し、図表示用クエリを生成して WA API に  
 送信し、返信結果として受け取ることでできる画像  
 を数式与式とともに表示することで、学習項目と図  
 形を同時に確認できるようにした。

#### 5. 成果の具体例

3.1 の切り替え規則ならびに正規化処理の適用に  
 より、項順の違い、冗長な指数表現、展開/因数分  
 解の違い、左右入れ替え、同類項整理、約分といっ  
 た差異を吸収できた。例として、 $y = xx^2 + bx + c$  を  
 検索すると、放物線のパターンと一致して、誤解の  
 学習項目が抽出されてしまっていた。これに対し、  
 WA API を用いて数式を正規化することで、以下の図  
 1のように、 $y = x^3 + bx + c$  に自動変換され、正し  
 く三次関数として抽出できた。

その結果、正規化前には判定できなかった式の一  
 部について、正規化後には既存の判定パターンと一  
 致し、学習項目の分類が安定することを確認した。

$y = xx^2 + bx + c$	$y = x^3 + bx + c$	$y = x^3 + bx + c$ 抽出した学習項目: 三次関数方程式
$y = ax^3x^2 + bx + c$	$y = ax^5 + bx + c$	$y = ax^5 + bx + c$ 抽出した学習項目: 不明方程式

図 1. WA API による正規化処理の結果

#### 6. 学習項目抽出とグラフ表示の検出率調査

本研究では、提案手法による学習項目抽出および  
 図表示機能の有効性を評価するため、数学教科書(数  
 学 I~数学 C)に基づき、図形分野に属する 67 種類  
 の方程式パターンをクエリとして事前登録した。

また、入力数式における表記揺れへの対応能力を  
 検証する目的で、既存の公式に対して以下の 5 種類  
 の表記変換ルールを適用、さらに派生数式を生成し  
 た。「登録済みの元の式」、「左辺・右辺の入れ替え」、  
 「文字係数を含む多項式」、「冗長な指数表現を含む  
 式」と「項の並び順が異なる式」である。

評価対象の中で直線を対象として、上記ルールに  
 より生成した派生式について、WA API による正規

化導入前後の検索結果を比較した。

その結果、学習項目抽出においては、25 件の評価  
 対象のうち、WA API による正規化を導入すること  
 で新たに 7 件の式が判定可能となり、検出率は 40%

(=10/25) から 68% (=17/25) に向上した。これは  
 数式の表記揺れが、WA API の正規化処理によっ  
 て吸収されたためである。また、グラフ表示に関し  
 て、84% (=21/25) について正常にグラフ生成が可能に  
 なった。正しく行えなかったケースは係数が文字で  
 あったために表示するための数値条件が網羅されて  
 いないがゆえに WA API が Plot を実行できなかった  
 ためであり、本質的な課題点には当たらない。

#### 7. まとめ

本研究では、数式検索システム [2]の学習項目  
 抽出機能 [2,3] と WA API を連携させ、図形カテ  
 ゴリの自動判定および対応グラフの表示に関する機能  
 拡張を行った。WA API から得られた正規化結果  
 (MathML) を比較表現として用いることで、項順  
 差・展開差・左右入れ替え・約分等の表記ゆれを吸  
 収し、分類の安定性を向上させることに成功した。

しかし、「左右の式が入れ替わっているケース」や  
 「項の並び順が正しくないケース」では、WA API で  
 正規化しても、数式の表現方法が多様であるため、  
 既存のパターンでは対応できないケースがまだ存在  
 する。今後は数式構造に WA API 正規化クエリを追  
 加するなどの検討も必要と考えられる。

#### 参考文献

- [1] W3C, “MathML 3.0,”  
<http://www.w3.org/TR/MathML3/> (参照 2026-01-27)
- [2] 渡部 孝幸, 宮崎 佳典, 正規表現を用いた数式検  
 索手法の提案, 情報処理学会論文誌, **56** (5), pp. 1417–  
 1427, 2015. <http://id.nii.ac.jp/1001/00142020> (参照  
 2026-01-27)
- [3] Y. Miyazaki, K. Shinshi, “Searching Mathematical  
 Expressions with Regular Expressions Tool and Its  
 Application to Extract Mathematical Concepts,” In: SITE  
 2017, pp. 1974–1978, 2017.
- [4] “Wolfram|Alpha,” <https://www.wolframalpha.com/>  
 (参照 2026-01-27)
- [5] Keisuke Yokoi, Akiko Aizawa, “An Approach to  
 Similarity Search for Mathematical Expressions using  
 MathML,” In: Towards a Digital Mathematics Library,  
 Masaryk University Press, pp. 27–35, 2009. [https://dml.  
 cz/handle/10338.dmlcz/702557](https://dml.cz/handle/10338.dmlcz/702557) (参照 2026-01-27)
- [6] 横井 啓介, Minh-Quoc NGHIEM, 松林 優一郎,  
 相澤 彰子, 意味と構造を考慮した数式検索手法の  
 提案, DEIM Forum 2011, B2-2, 2011. [https://db-  
 event.jp.org/deim2011/proceedings/pdf/b2-2.pdf](https://db-event.jp.org/deim2011/proceedings/pdf/b2-2.pdf) (参照  
 2026-01-27)
- [7] 脇 弘太, 宮崎 佳典, 代数的変形に対応した変形  
 依拠公式提示ツールの開発, 情報処理学会 第 82 回  
 全国大会, pp. (4)-747–748, 2020.