

数式変形依拠公式提示アプリケーションの機能拡張への試み An Extension to the Function of Extracting Math Concepts and Formulas in Transforming Mathematical Expressions with Annotation

加藤 駿弥^{*1}, 宮崎 佳典^{*2}, 中村 泰之^{*3}, 田中 省作^{*4}, 新谷 誠^{*2}
Shunya KATO^{*1}, Yoshinori MIYAZAKI^{*2}, Yasuyuki NAKAMURA^{*3}, Shosaku TANAKA^{*4}, Makoto ARAYA^{*2}

^{*1} 静岡大学 情報学部

^{*1} Faculty of Informatics, Shizuoka University

^{*2} 静岡大学学術院 情報学領域

^{*2} College of Informatics, Shizuoka University

^{*3} 名古屋大学 大学院情報学研究科

^{*3} Graduate School of Informatics, Nagoya University

^{*4} 立命館大学 文学部

^{*4} College of Letters, Ritsumeikan University

Email: kato.shunya.18@shizuoka.ac.jp

あらまし：我々は自己開発した数式検索システムの応用として、変形されていく数式に対し、公式適用箇所をその公式名とともにアノテートする“変形依拠公式提示機能”を開発している。しかし、事前登録した公式に代数的変形が施された上で適用されている場合は抽出できず、数式処理システム（CAS）の Wolfram Alpha（API）を利用して適用範囲を広げた先行研究があるが、その範囲は限定的である。本研究では他の CAS を対象にカバー率の高い同機能への拡張を目指す。

キーワード：数式検索システム、正規表現、変形依拠公式提示機能、Wolfram Alpha API、SymPy

1. はじめに

STEM 教育ならびに数理・データサイエンス教育が世界の潮流となってきた。しかし、それらの教育において必要不可欠な数式に対し、苦手意識を抱く学習者は多く存在する。一方で、MathML などを代表に計算機上で数式を利用するためのデータ形式の普及が進み、STEM 教育で利用されるデジタル教科書の普及などが今後予想される。そこで我々研究グループは、計算機上の数式を含む文書を対象とした数式検索システムを開発し⁽¹⁾、将来の数学学習に役立てる機能拡張を行ってきた。さらにはその機能の 1 つである、数式変形の拠り所となっている公式を検索・提示する変形依拠公式提示機能を数式処理システムと連携することで拡張し、数学学習支援においてより高い有用性を持つ数式検索システムの構築を目指した⁽²⁾。本発表では⁽²⁾が Wolfram Alpha（API）と連携したのに対し、さらに SymPy を適用し、両者を相補的に用いることでさらなるカバー率向上を目指すことを目的とする。

関連研究として⁽³⁾は、数式処理システム Mathematica や動的幾何学ソフト GeoGebra を用いた中等教育数学向けソフトウェア教材を作成した。⁽⁴⁾は CERN Document Server（CDS）における数学コンテンツの問い合わせを行う検索システムを開発した。

2. 変形依拠公式提示機能

Web 上の文書に数式を記述する場合に数式表示用のデータ形式が用いられる。そこで⁽¹⁾のシステムは、MathML 内の表示の情報を記述するタグセットである Presentation Markup で記述された（以下、“MathML 記述”と表記）数式を対象とし、正規表現とパター

ンマッチングによるページ内数式検索を基本的な機能として実装した。文書中の MathML と検索窓に入力した検索数式を正規表現へ置換ののち照合し、マッチした文書中当該箇所の数式をハイライトする。

数式検索システムの機能として実装されたのが変形依拠公式提示機能である。機能の適用例として図 1 を示す。文書中の数式変形を読み込み、公式クエリ（あらかじめシステムに登録されている MathML 記述の公式データ）と照合することで、式変形の拠り所となっている公式名を提示する。



図 1 変形依拠公式提示機能

3. Wolfram Alpha API

⁽²⁾は変形依拠公式提示機能に Wolfram Alpha API（以下 Wolfram Alpha）と連携することで機能を拡張し、公式クエリから代数的に変形している数式に対応できるようにした。Wolfram Alpha に MathML

記述の数式を渡すとその変形結果として MathML 記述の数式が返される。実験では高校数学レベルで学ぶ 103 個の公式に対して次のルールを適用した 5 種の派生数式を用意した。

- ・ 従前の登録済みと同一
- ・ 両辺を入れ替え
- ・ 両辺×N
- ・ 両辺÷M
- ・ 両辺×(-1)

(ここに N, M は適当な自然数)。“両辺×N”、“両辺÷M”についてはルールを適用した数式の MathML 木構造が崩れる場合のみ適用する。ルールを適用した数式を Wolfram Alpha に渡すと一定数共通の数式に変形されて返される。5 種の数式の変形結果がすべて一致するもの、2 式や 3 式、4 式一致するものがある。この共通の数式を基底式として公式クエリに登録した。実際に 5 種のルールにより得られた数式は 283 個である。基底式を得られた公式は 103 個中 83 個あり、削減できた数式は 149 個だった。

[削減できた数式 (149 個)] / [従前の登録済み公式を除いた派生数式 (283-103=180 個)] として削減率を求め、約 82.8% を削減することに成功した。283 個必要だった公式クエリが大幅に削減できている。基底式が取得できたことで公式クエリの登録数を増やしつつ、幅広い数式に対して公式提示が実現されている。

4. 変形依拠公式提示機能の改良案

Wolfram Alpha と連携した変形依拠公式提示機能でも対応できていない公式が存在していることに対し、本研究ではそれに代わる、より高精度の数式処理システムを模索した。その結果、SymPy⁽⁵⁾を利用すると数式の単純化がなされ、共通な式(基底式)が得られる可能性のあるものが多く見つかった。そこで(2)と同様に公式数 103 個の数式に対して 5 種の代数的に変形させて得た数式に対して単純化の処理を試み、その削減率を調べた。SymPy の機能として Simplify が用意されており、単純化を施すことができる。103 個の公式から代数的変形で得た 180 個の数式のうち共通な式に置き換えることができた数式は 100 個となった。削減率は 55.6% だった。Wolfram Alpha 適用時の削減率は約 82.8% であったことを考えると大きく劣る性能であったことがわかった。代替数式処理システムも探したがこれ以上のものは見つからなかった。そこで本研究では Wolfram Alpha で未だ対応できていない公式に対して SymPy により基底式を得るケースの割合について調査した。Wolfram Alpha と SymPy を相補的に利用することで削減率を向上させ、対応できる公式数を増やすことで変形依拠公式機能の実用化を目指せるからである。

Wolfram Alpha で対応できていない公式の代表例としてコーシー・シュワルツの不等式 (2 次) について単純化を行った例を示す。同不等式の従前の登録されている公式クエリは以下である。

$$(a_1^2 + a_2^2)(b_1^2 + b_2^2) \geq (a_1a_2 + b_1b_2)^2$$

コーシー・シュワルツの不等式 (2 次) は 5 変形のうち、以下のルールで変形したものが対象となる:

- ・ 従前の登録済みと同一
- ・ 両辺を入れ替え
- ・ 両辺×(-1)

以上のルールを適用した数式に対し、単純化を行った結果、以下の共通の数式を得られた。

$$(a_1a_2 + b_1b_2)^2 \leq (a_1^2 + a_2^2)(b_1^2 + b_2^2)$$

以上は 1 例であるが、Wolfram Alpha 単独では対応できない公式は 26 種類、その内ルールを適用して得た式は 31 である。この 31 個の数式について同様に処理を行った結果、共通な数式に変形でき、かつ削減できた数式は 13 個であった。Wolfram Alpha で削減できた 149 個と合わせると 162 個削減できたことになる。このときの削減率を計算すると、[削減できた数式 (162 個)] / [従前の登録済み公式を除いた派生数式 (180 個)] により 90.0% に向上した。

コーシー・シュワルツの不等式に関しては、5 種の派生形式に追加で両辺を入れ替えた場合と両辺×(-1)を同時に適用した数式についても試行した結果、同等の基底式が得られた。現時点で派生数式は上述 5 種の代数的変形のみであるが、今後は他の変形についても考究してゆく必要があると思われる。

5. まとめ・今後の展望

本研究では、代数的に変形させた公式に対して共通の数式に変形させることで変形依拠公式機能のカバー率を上げることを目的として、今回 SymPy で数式変形を行い、Wolfram Alpha で得られなかった基底式を得ることが確認された。相補的に使用することを考慮すれば、カバー率すなわち削減率が向上したと認めることができる。一方で SymPy を利用するには問題点も残っており、入力として受け付ける LaTeX 形式から MathML 形式への変換処理などが挙げられる。出力形式は MathML 形式が指定できるため問題はない。今後の展望としては、SymPy を変形依拠公式機能に実装していくことで実用化を目指していくことである。

参考文献

- (1) 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, Vol.56, No.5, pp.1417-1427 (2015)
- (2) 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)
- (3) 佐々木 重雄, 数式処理システム Mathematica と動的幾何学ソフト GeoGebra を用いた中等教育向け教材集の試作, 秋田大学教育文化学部研究紀要, 教育科学部門, 73, pp. 19-26 (2018)
- (4) Arthur Oviedo, 5e^(x+y): A Math Aware Search Engine (for CDS), CERN-THESIS-2014-012 (2014)
- (5) SymPy :<https://docs.sympy.org/latest/modules/simplify/simplify.html>