

変形依拠公式提示機能への拡張も含めた数式検索システムの応用

野崎 咲†

宮崎 佳典‡

田中 省作*

静岡大学情報学部†

静岡大学大学院情報学領域‡

立命館大学文学部*

1. 導入

計算機上で数式を扱うためのデータ形式が普及し、それに伴い数式データが蓄積されたことによって、数式を検索するための技術の必要性が高まってきている。本研究室では、先行研究にて MathML Presentation Markup 形式で記述された数式を対象とした数式検索システムの開発を行い、またこのシステムを応用した機能の追加を行ってきた[1]。本発表では、その応用例として、式変形の拠り所となっている公式を提示する変形依拠公式提示機能の拡張、および数式検索システムを用いた際の検索履歴データを蓄積、活用するための機能の実装について述べる。これらの機能追加・拡張により、数式検索システムの技術を応用した、数学教育の学習支援に役立つシステム構築を目指す。

関連研究の例として、Nidhin ら[2]は MathML Presentation Markup で記述されている数式の類似度を算出し、これを用いて検索を行うシステムの開発を行っている。山田ら[3]は数学用語をクエリとして検索を行い、web ページからクエリに関連する画像を抽出するシステムを開発している。また澁谷らの研究[4]では、MathML 形式の数式の類似度を用いた検索を、数式単位ではなく、数式の部分構造に対して行うことで、対象数式を理解する上で有用な数式のアノテーションを可能にするシステムの開発を行っている。

2. 開発システムの概要

2.1 数式検索機能

MathML 記述の数式に対し、パターンマッチングにより数式検索を行うもので、見た目の数式は同じでも複数の MathML Presentation Markup 記述が存在することを鑑み、正規化処理を事前に行う。クエリに対し、一致した数式の部分がハイライトされる。柔軟な検索を実現するため、図 1 の左側の表に示すような正規表現も入力することができる。図 1 のクエリ例は、後方参照機能を用いることで同一の数式同士の乗算（図 1 の右上）を検索している。このクエリ表現によ

り、掛け算の表記が省略されている場合 (bb) も乗算記号が付してある場合 ($b \times b$ や $b \cdot b$ など) も同時にヒットさせることができる。

	内部入力文字のいずれか	
	内部の文字列と最大1回マッチ	
	仕切りごとの入力文字列のいずれか	

図 1 数式検索の適用例

2.2 変形依拠公式提示機能

変形依拠公式提示機能とは、公式データベース(数式検索システムを用いて作成した公式のクエリを登録したもの)の各公式のクエリと照合することで、式変形の拠り所となっている公式や公式名を提示する機能である。図 2 に例を示す点線で囲まれた数式に対して実行すると、実線で囲まれた数式のように、公式名の表示に加えて、式変形に用いられた公式適用箇所がハイライトされる。点線で囲まれた数式の 1→2 行目への変形が実線で囲まれた数式の 2 行目と 3 行目に、点線で囲まれた数式の 2→3 行目への変形が実線で囲まれた数式の 4 行目と 5 行目に示されている。

$\sin(x + \pi) = \sin\left(\left(x + \frac{\pi}{2}\right) + \frac{\pi}{2}\right)$	
$= \cos\left(x + \frac{\pi}{2}\right)$	
$= -\sin x$	

図 2 変形依拠公式提示機能の適用例

3. 変形依拠公式提示機能の改良

[1]に現時点で実装されている変形依拠公式提示機能には、データベースに登録されているクエリに完全一致する式変形のみヒットするという強い制約がある。相加相乗平均を用いた変形

Development of a Tool Annotating Underlying Math Formula as Application of Math IR System

†Saki Nozaki, Faculty of Informatics, Shizuoka University

‡Yoshinori Miyazaki, College of Informatics, Shizuoka University

*Shosaku Tanaka, College of Letters, Ritsumeikan University

を例示する。データベースには、図 3 のクエリが登録されている。[1]のシステムでは、図 3 のクエリと数式が一致する図 4(1)のような数式には公式提示が付されるが、 c が両辺に加算された図 4(2)の数式にはマッチしない。一方で、一般的な式変形で頻出するのは後者のような、与式の一部が変形されているケースである。この問題に対し、変形の前後の同一箇所のみを両辺から取り除く差分抽出により解決する。

図 3 相加相乗平均の公式用数式クエリ

(1) $\frac{a+b}{2} \geq \sqrt{ab}$ ← 相加相乗平均 $\left(\frac{a+b}{2} \geq \sqrt{ab}\right)$
 (2) $\frac{a+b}{2} + c \geq \sqrt{ab} + c$

図 4 図 3 の数式クエリによる適用結果例

たとえば、図 4(2)の木表現は簡略的に表記すれば図 5 のようになる。両辺の木をたどっていくと、左辺と右辺で一致しているのは右端の c であることがわかる。両辺から一致した要素を取り除いたものが図 6 であり、これは数式 $\frac{a+b}{2} \geq \sqrt{ab}$ を表し、データベースに登録されているクエリ(図 3)にマッチする。その後、削除した要素を復元すると、図 6 の下部に示したように、変形に使用された公式部分がハイライトされ、公式名が提示される。

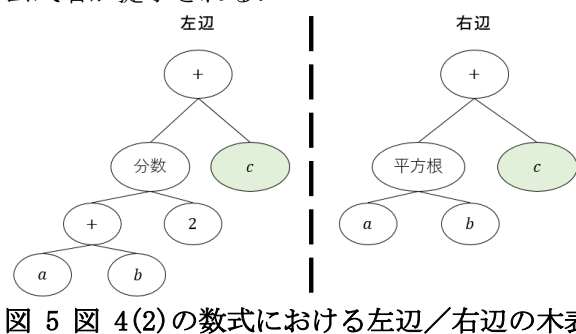


図 5 図 4(2)の数式における左辺/右辺の木表現

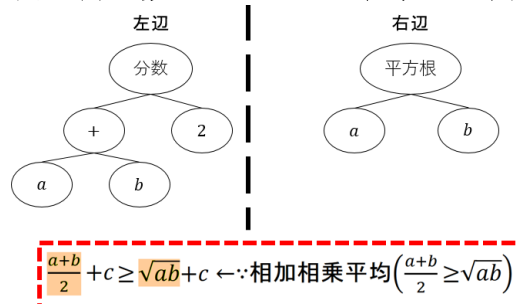


図 6 左辺/右辺の木表現の差分抽出結果

4. 検索履歴機能

検索履歴機能は、ユーザの利便性を考慮し、数式検索機能による検索履歴を保存・蓄積する機能である。数式検索のたびに入力クエリが保存され、検索履歴欄に表示される。過去に入力されたクエリは、コピー&ペーストで入力欄に再度入力ならびに編集することができる。これにより、同一あるいは類似した数式を検索する場合の作業効率を高めることができる。

5. まとめと今後の展望

本研究では、数式検索システム[1]の応用として位置付けられている変形依拠公式提示機能の強化を行い、数式変形中の変形箇所のみを抽出したものを、データベースに登録されているクエリに照合できるように改良した。また検索履歴機能の実装を行い、ユーザの利便性を向上させた。

今後の展望として、変形依拠公式提示機能の更なる改善が第一に挙げられる。データベースに登録されている公式数の拡充や公式検索に要する処理時間の短縮により実用化を目指す。さらに、公式データベースに登録されている数式から少しでも代数的に変形したもの(例:公式の両辺を2で割ったものや、右辺を左辺に移行したもの)にはヒットしない現状など、未解決の課題にも取り組んでいく。また、ユーザのログイン機構を備えることで、利用履歴からユーザの苦手分野・興味の特長などのパーソナライゼーションにも繋がりたいと考えており、ソーシャルラーニングなどの新しいスタイルによる学習支援も視野に入れたい。

参考文献

- [1]. 宮崎 佳典, 粥川 佳哉, 田中 省作, 正規表現機能付き数式検索手法ならびに学習項目抽出への応用, 電子情報通信学会技術研究報告 信学技報 117(83), pp. 41-44 (2017)
- [2]. Pattaniyil Nidhin, and Richard Zanibbi, Combining TF-IDF Text Retrieval with an Inverted Index over Symbol Pairs in Math Expressions: The Tangent Math Search Engine at NTCIR 2014, NTCIR, pp. 135-142 (2014)
- [3]. 山田 奉子, et al., 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, Vol. 33, No. 4, pp. A-H91_1-13 (2018)
- [4]. 澁谷 海渡, et al., 部分構造を考慮した数式理解支援システムの提案, 人工知能学会研究会, SIG-KBS-B501-01, 6 pages (2015)