

MathML を用いた類似数式検索方式の実現

岸本 貞弥[†] 中西 崇文^{††} 櫻井 鉄也^{†††} 北川 高嗣^{†††} 栃木 敏子^{††††}

[†] 筑波大学 第3学群 情報学類 〒305-0006 茨城県つくば市天王台 1-1-1 数値解析研究室

^{††} 筑波大学大学院 システム情報工学研究科 〒305-0006 茨城県つくば市天王台 1-1-1 数値解析研究室

^{†††} 筑波大学 電子・情報工学系 〒305-0006 茨城県つくば市天王台 1-1-1

^{††††} 東京大学 情報基盤センター 〒113-8658 東京都文京区弥生 2-11-16

E-mail: [†]kishimoto@nalab.is.tsukuba.ac.jp, ^{††}takafumi@nalab.is.tsukuba.ac.jp,

^{†††}{sakurai,takashi}@is.tsukuba.ac.jp, ^{††††}tochigi@cc.u-tokyo.ac.jp

あらまし 本稿は, MathML(Mathematical Markup Language) で記述された数式を対象とした類似数式検索の実現方式について示す. 現在, 多くの数式を含んだ技術ドキュメントが web 上に散在している. これらの文書中における数式の大半は gif 形式などの画像に変換されてドキュメントに貼り付けられているのが現状である. そのため, 数式の意味を自動的に抽出することが困難であり, 数式を対象とした検索方式は実現されていない. 一方, MathML により, 数式の表示と内容の記述が技術的に可能となり, 各種数式ソフトに対応されつつある. 本稿では, MathML で記述された数式を対象として, MathML のタグ情報に注目したデータについて Latent Semantic Indexing を適用することにより, 数式検索を行う方式について示す. 本方式により, 数式を構成している演算子に注目することで, MathML で表された数式を問い合わせとした類似数式の検索を可能としている.

キーワード MathML, 類似数式検索, Latent Semantic Indexing, tf · idf

An Implementation Method of Similarity-Based Retrieval for Formulas using MathML

Sadaya KISHIMOTO[†], Takafumi NAKANISHI^{††}, Tetsuya SAKURAI^{†††}, Takashi KITAGAWA^{†††},

and Toshiko TOCHIGI^{††††}

[†] College of Information Sciences, Third Cluster of Colleges, University of Tsukuba

^{††} Graduate School of Systems and Information Engineering, University of Tsukuba

^{†††} Institute of Information Sciences and Electronics, University of Tsukuba

^{††††} Information Technology Center, the University of Tokyo

E-mail: [†]kishimoto@nalab.is.tsukuba.ac.jp, ^{††}takafumi@nalab.is.tsukuba.ac.jp,

^{†††}{sakurai,takashi}@is.tsukuba.ac.jp, ^{††††}tochigi@cc.u-tokyo.ac.jp

Abstract This paper presents an implementation method of similarity-based retrieval for mathematical formulas described in MathML. Many technical-writings documents containing mathematical formulas are scattered on the web. Almost all the mathematical formulas in these documents are transformed into the picture file. Since it is difficult to extract the meaning of mathematical formulas automatically, similarity-based retrieval system for mathematical formulas is not realized. On the other hand, the presentation of mathematical formulas and description of the mathematical contents become possible technically by MathML, which is widely adopted in various softwares.

This paper presents an implementation method of similarity-based retrieval for mathematical formulas described in MathML by using Latent Semantic Indexing taking the tag information of MathML into consideration. This method enables acquisition of the similar mathematical formulas by using MathML expression as the query.

Key words MathML, Similarity-based formulas retrieval, Latent Semantic Indexing, tf · idf

1. はじめに

現在、様々なドキュメントがデジタル化されて広域ネットワーク上に散在している。また、ドキュメント数は増加を続けており、それらのデータ群は、知識・情報の源として重要な存在となっている。このような環境下で、これらの大規模なドキュメント群を対象とし、利用者の要求に適合した知識、情報を提供可能な性能の高い検索手法が必要とされている。

ドキュメントを扱う検索方式として、本文の内容によって検索を行うシステムは実現されている。しかしながら、科学論文などにおいては数式が多く含まれており、それらの数式の意味が重要となる場合が多い。このような科学論文等の数式を含んだドキュメントについて、意味的な内容を反映した検索を行うためには、本文の内容だけではなく、数式の内容を端的に表すメタデータをドキュメントごとに付与する必要がある。また、そのメタデータを対象とした数式の内容による計量方式の実現が重要があると考えられる。

これまで、ドキュメント中に含まれる数式の大半は CompuServe Graphics Interchange Format (gif) 形式などの画像に変換され、本文中に貼り付けられるのがほとんどであった [1]。画像データは文書のデータを大きくするだけでなく、修正も面倒である。その上、画像データにすることにより数式の意味的な情報は失われ、メタデータの自動抽出など、数式の内容に対応した自動的な処理が困難である。

一方、1998年4月に World Wide Web Consortium (W3C) の数式ワーキンググループにより、Extensible Markup Language (XML) の応用技術である Mathematical Markup Language (MathML) [2] Version 1.0 が公表され、web 上での数式の表示と内容の記述が可能になった。2001年2月には MathML Version 2.0 が公表され、各種 web ブラウザや数式処理ソフトも対応しつつある [3]。例えば、数式処理ソフトである Maple、Mathematica は、それぞれ独自の数式表記から MathML 表記に変換できる。また、TeX で書かれたドキュメントを MathML を含む文書に変換するソフトウェアもある [4]。

このように、数式を MathML で容易に表記できる環境が整えば、ドキュメント中に含まれる数式を対象として、数式の内容を端的に表すメタデータの自動抽出が可能となり、数式を対象とした検索が実現できると考えられる。

本稿では、MathML で表記された数式を対象とした類似数式検索方式について示す。本方式は MathML のタグの出現頻度から抽出されたメタデータに対して Latent Semantic Indexing [5] を適用することにより、数式検索を行う方式である。MathML のタグは、数式の演算子などを表しているため、このタグに注目してメタデータを抽出することにより、数式に使われている演算子や記号による数式の内容の類似検索が可能となる。これは同じ公式であってもドキュメントによって添え字や変数に使う文字に違いが生じることがあるが、タグ情報に注目することで同じ公式として検索可能であることを意味している。

今後、数式の検索と全文検索を組み合わせることにより、数式を含む技術文書を対象としたドキュメント検索が実現でき

```
<math>
<mrow>
<msup>
<mfenced>
<mrow>
<mi>a</mi>
<mo>+</mo>
<mi>b</mi>
</mrow>
</mfenced>
<mn>2</mn>
</msup>
</mrow>
</math>
```

図 1 $(a+b)^2$ の Presentation Markup による表記。

Fig. 1 Notation by Presentation Markup of $(a+b)^2$.

ると考えられる。これにより、利用者の要求により合致したドキュメント検索方式が実現されることが考えられる。

本稿では、MathML で表された数式を対象とし、数式を問い合わせとした類似数式検索方式について示す。また、実験により本方式の有効性を検証する。

2. Mathematical Markup Language (MathML)

MathML は数式の構造と内容の両方を書き表すことを可能とする XML ベースのマークアップ言語である。MathML ファイルは単独で使用されるほか、他の XML 文書に埋め込んで使用することが可能である。MathML は特に XHTML で記述された Web ページに数式を埋めこむ際に使われることを強く意識されている。MathML 文書は Maple などの数式処理アプリケーションで扱うことができるほか、XHTML との連携を意識して、Web ブラウザでも対応が進んでおり、Mozilla や Netscape7 では既に対応している。Internet Explorer では、プラグインの MathPlayer を使用することで MathML に対応できる。

MathML では数式の表記を表す Presentation Markup と数式の意味を扱う Content Markup の 2 種類のタグが用意されている。以下でそれぞれについて説明する。

• Presentation Markup

Presentation Markup は数式の持つ意味は表現しておらず、少数のタグセットから成る。これは主に web ブラウザなどでの数式表示を行うために用いられる。Presentation Markup で表現された数式例を図 1 に示す。

• Content Markup

Content Markup は省略された積など数式の表示には明確に現れてこない構造も含め、数式の内容を正確に記述するためのものであり、約 150 個のタグが用意されている。Content Markup で表現された数式例を図 2 に示す。

本稿では、数式の内容を記述する Content Markup を検索に利用する。

```

<math>
  <apply>
    <power/>
  </apply>
  <plus/>
  <ci>a</ci>
  <ci>b</ci>
  </apply>
  <cn>2</cn>
  </apply>
</math>

```

図2 $(a+b)^2$ の Content Markup による表記 .
Fig.2 Notation by Content Markup of $(a+b)^2$.

3. Latent Semantic Indexing

本節では、問い合わせと検索対象が共通の単語 (属性) を持たなくても検索可能な方式である Latent Semantic Indexing (LSI) [5] について示す。LSI は、特異値分解を用いて用語と文書間の関係行列を直交行列に分解し、行列の次元数を減らすことにより、共起性の高い語を 1 つに縮退する方法である。

3.1 特異値分解

特異値分解とは、データ行列 \tilde{D} が与えられたとき、

$$\tilde{D} = U_r \Sigma_r V_r^T \quad (1)$$

のように 3 つの行列 U_r, V_r, Σ_r に分解することである。ただし、 $U_r \in \mathbb{R}^{m \times r}$ 及び $V_r \in \mathbb{R}^{n \times r}$ は列正規直交行列である。すなわち、 E を単位行列としたとき、

$$U_r^T U_r = V_r^T V_r = E \quad (2)$$

を満たす行列である。また、 $\Sigma_r \in \mathbb{R}^{r \times r}$ は

$$\Sigma_r = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \quad (3)$$

となり、 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > \sigma_{k+1} = \sigma_{k+2} = \dots = \sigma_r = 0$ を満たす。

ここで、 $\sigma_1, \sigma_2, \dots, \sigma_r$ は \tilde{D} の特異値という。この関係から、 σ_1 に比べて十分小さい値を ϵ として、 $\sigma_k > \epsilon > \sigma_{k+1}$ ($k \leq r-1$) となるとき、 \tilde{D} は

$$\tilde{D} \simeq D_k = U_k \Sigma_k V_k^T \quad (4)$$

と、近似できる。但し、 $U_k \in \mathbb{R}^{m \times k}$ 、 $V_k \in \mathbb{R}^{n \times k}$ 、 $\Sigma_k \in \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k)$ である。

3.2 問い合わせと検索対象データとの計量方式

本節では、ユーザからの問い合わせと検索対象データとの比較方法について示す。

3.2.1 検索対象データのベクトル表現

初めに、 n 個の検索対象データ d_i ($i = 1, 2, \dots, n$) を m 個の特徴 f_1, f_2, \dots, f_m で以下のように特徴づけされたメタデータを用意する。

$$d_i = (f_{1i}, f_{2i}, \dots, f_{mi})^T. \quad (5)$$

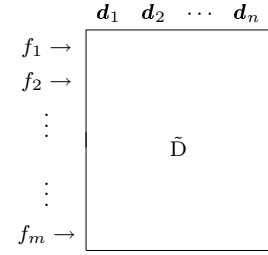


図3 データ行列 \tilde{D} .
Fig.3 data matrix \tilde{D} .

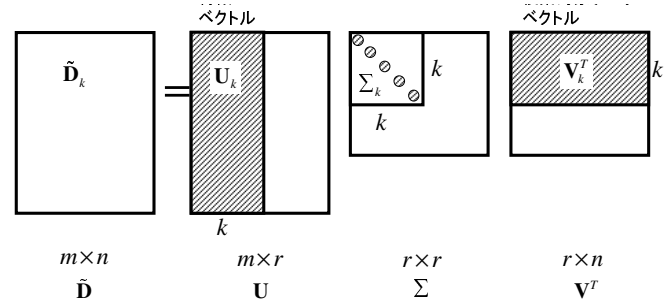


図4 D_k の模式図.
Fig.4 The figure of D_k .

そのベクトルを並べて構成する図 3 のような $m \times n$ 行列をデータ行列 \tilde{D} とおく。

さらに、式 (4) により、データ行列 \tilde{D} を図 4 のように特異値分解を行う。ここで、 D_k を以下のように定義する。

$$D_k = [d_{k1}, d_{k2}, \dots, d_{kn}]. \quad (6)$$

これにより、特徴及び検索対象データをそれぞれ、空間内にベクトル表現することができる。

3.2.2 問い合わせのデータ表現と計量方式

ユーザから発行される問い合わせは、データ行列 \tilde{D} を作る際に使用した m 個の特徴で次のように表される。

$$q = (q_1, q_2, \dots, q_m)^T. \quad (7)$$

ここで、 q の各成分 q_1, q_2, \dots, q_m は、それぞれ f_1, f_2, \dots, f_m に対応している。さらに、 q を k 次元ベクトル表現したものを

$$\hat{q} = (\hat{q}_1, \hat{q}_2, \dots, \hat{q}_k)^T \quad (8)$$

とする。 \hat{q} は、

$$\hat{q}^T = q^T U_k \Sigma_k^{-1} \quad (9)$$

で求められる。これにより問い合わせを k 次元空間にベクトル表現でき、既に k 次元空間にベクトル表現している検索対象データと比較が可能となる。問い合わせと検索対象データとの余弦 (cos) の値によって類似度を計量する。具体的には以下のように表される。

$$\cos \theta_{ki} = \frac{(\hat{q}, d_{ki})}{|\hat{q}| |d_{ki}|}. \quad (10)$$

$\cos \theta_{ki}$ の値がある閾値を上回った検索対象データについて、値が大きい順にソートされ出力される。

4. 類似数式検索の実現方式

本節は、MathML で書かれた数式を対象として、数式で与えられる問い合わせによる類似数式検索方式について示す。

4.1 類似数式検索方式の概要

ここでは、類似数式検索の実現方式の概要を述べる。本方式は、MathML で書かれた数式を対象として、与えられた数式と近い数式を検索するシステムである。本方式の特徴は、数式の演算子に注目して検索を行うことにより、添え字や変数に使う文字の違いなどによる、記述方法が異なる数式においても同様の意味と捉えて検索可能な点にある。本方式の概要を以下に示す。

(1) 検索対象の数式データ群よりデータ行列自動作成

まず、検索対象の MathML で記述された数式から、その数式の特徴を表すメタデータを抽出する。次にそれらを並べて構成するデータ行列を生成する。これにより、検索対象の数式データ群を類似度を測定する空間に表現することができる。メタデータ自動抽出方式については 4.2 節で示す。

(2) 問い合わせの数式よりメタデータ抽出

検索対象の数式データと同様に、問い合わせとして与えられた MathML で記述された数式から、その数式の特徴を表すメタデータを抽出する。

(3) 類似度測定

上記項目 (1),(2) により抽出されたメタデータから、式 (10) によって類似度を計量し、その値の大きい順にソートする。これにより、問い合わせと合致した数式が検索される。

4.2 MathML で表現された数式を対象としたメタデータ自動抽出方式

本節では、MathML で記述された数式からメタデータを抽出する方式について述べる。本方式は、MathML のタグ情報に注目し、数式の特徴として抽出することにより、数式の演算子に依存した検索を実現するものである。具体的には以下の手順で実現される。

(1) MathML 表現の数式が構成するタグの種類とその出現頻度を導出

対象となる MathML 表現の数式データ $p_i (i = 1, 2, \dots, n)$ のタグの種類とその出現数をカウントすることで特徴づけする。

$$p_i = (f_{1i}, f_{2i}, \dots, f_{mi})^T. \quad (11)$$

f_1, f_2, \dots, f_m は対応する MathML のタグの出現頻度を表す。例として図 5 のように行う。

(2) tf · idf による重み付け

抽出したタグの頻度によってその数式の特徴を表しているが、タグの中には、どの数式にも多く含まれるタグが存在し、各数式の特徴を表す際にノイズとなる可能性がある。

本方式では、全文検索においてよく用いられている tf · idf [7], [8] を用いて重み付けを行う。tf · idf は、1 データ内でより出現頻度の高い特徴についてより大きな値を示し、かつデータ群全体で、より出現頻度の大きい特徴についてより小さな値を示す。tf · idf の値をメタデータ生成に反映させることに

```

<math>
  <apply>
    <sin/>
    <ci>x</ci>
  </apply>
</math>

```

sin x の MathML 表現

	...	domain	apply	sin	cos	cn	ci	plus	...
sin x	...	0	1	1	0	0	1	0	...

MathML 中のタグの種類とその頻度をカウント

図 5 sin x の例。

Fig. 5 Example of sin x.

より、各数式データのより明確な特徴を抽出することが可能となると考えられる。これにより、問い合わせにおいて各数式の意味をより反映した検索が可能となる。

具体的には、全データ群において該当タグが出現する頻度を g_j とすると、重み w_{ij} を以下のように表す。

$$w_{ij} = f_{ij} \log \frac{n}{g_j}. \quad (12)$$

これらより、数式データ p_i のメタデータは以下ようになる。

$$p_i = (w_{1i}, w_{2i}, \dots, w_{mi})^T. \quad (13)$$

5. 実 験

本方式の有効性を検証するため、本方式に基づく実験システムを構築し、検証実験を行った。

5.1 実 験 1

5.1.1 実験環境

検索対象データとして、W3C の MathML Test Suite [9] から収集した 357 個の Content Markup で書かれた数式と、Maple(version7.00) を使って Content Markup で書き出した 23 個の数式の計 400 個を用いる。Maple を使って書き出した式は、数学公式・数表ハンドブック [10] から引用した。

検索対象データからのメタデータの抽出では、まず XML パーサを用いて DOM ツリーを生成した。この DOM ツリーからタグの出現頻度をカウントした。これらの処理では Java プログラムを用いた。また、特異値分解には Matlab(version 5.2) [11] を用いた。

図 6 に実験システムの概要を示す。このシステムにおいてサーバー側の処理は Java Servlet により行った。表 1 にはシステムに用いたアプリケーションを示す。

5.1.2 実験方法

4.2 節で示したメタデータ抽出方式において、(i)tf · idf を適用せず出現頻度による重み付けによる方法、(ii)tf · idf を適用して重み付けの方法について比較を行った。また、本類似数式検索の精度について評価を行った。

比較方法については、まず本方式の性質を考察するために、実際の上位 5 位検索結果を見比べた。さらに、検索結果を再現率、適合率という指標によって評価した。ここで再現率、適合

表 1 実験環境.

Table 1 Experiment environment.

WWW ブラウザ	
Microsoft Internet Explorer	version 6.0
プラグイン	
MathPlayer	version 1.0
使用言語	
Java™ 2 Platform, Standard Edition	version 1.4.0
XML パーサ	
Xerces	version 1.2.3
サーブレット環境	
Tomcat	version 3.2.4
Apache	version 1.3.27

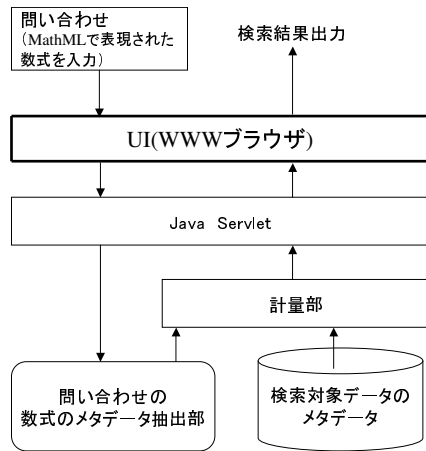


図 6 実験システム図.

Fig. 6 A figure of the system structure for experiments.

表 2 問い合わせと正解.

Table 2 queries and correct answers.

問い合わせ	正解データ	正解数
$\sin t$	$\sin x$ などの \sin を含む数式	33
$\max\{x, y\}$	$\max\{a, b\}$ などの \max を含む数式	3
$\lim_{x \rightarrow a} x$	$\lim(\sin x)$ などの \lim を含む数式	9
$\int_0^a (x+a)dx$	$\int_0^a (x+a)dx$ などの \int を含む数式	11

率は以下のように表される.

$$\text{再現率} = \frac{\text{システムの検索結果に含まれる正解数}}{\text{本来の正解数 (正解として出力されるべきデータ数)}} \quad (14)$$

$$\text{適合率} = \frac{\text{システムの検索結果に含まれる正解数}}{\text{システムの検索結果出力数}} \quad (15)$$

なお, 本実験で使用する問い合わせと正解を表 2 に示す.

5.1.3 実験結果

まず, 本方式の性質を考察するために, 問い合わせ $\sin t$, および $\max\{x, y\}$ について上位 5 位の検索結果をそれぞれ図 3, 図 4 に示す.

表 3 実験結果 1.

Table 3 Experiment results1.

問い合わせ : $\sin t$		
順位	(i)tf · idf を適用しない	(ii)tf · idf を適用
1	$\sin x$	$\sin x$
2	$\sin(a + b)$	$\sin(a + b)$
3	$\sin(x(y + z)z)$	$\sin(x(y + z)z)$
4	$\sin^{-1}(x)$	$\frac{1}{\sin t}$
5	$\sin x + \cos x$	$\sin x + \cos x$

表 4 実験結果 2.

Table 4 Experiment results2.

問い合わせ : $\max\{x, y\}$		
順位	(i)tf · idf を適用しない	(ii)tf · idf を適用
1	$\max\{a, b\}$	$\max\{a, b\}$
2	$\max\{x x \in B \wedge x \notin C\}$	$\max\{x x \in B \wedge x \notin C\}$
3	$(x, y) \rightarrow (f(x, y), g(x, y))$	$\max\{x - \sin x x > 0 \wedge x < 1\}$
4	$A \times B = ab \sin \theta N$	x
5	$\sum_{x=a}^b f(x) \sum_{x \in B} f(x)$	$x + x + x + y + \theta + \nu + \dots$

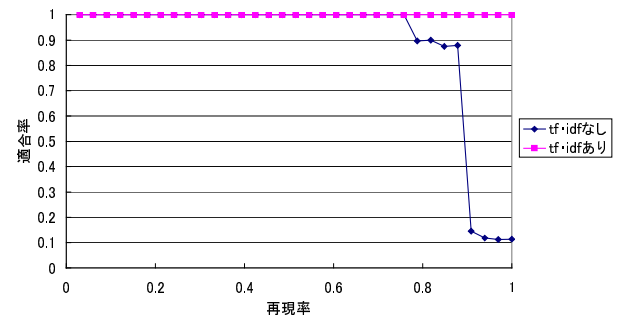


図 7 問い合わせ $\sin t$ の場合の再現率 · 適合率.

Fig. 7 Recall rates and precision rates for the query $\sin t$.

問い合わせ $\sin t$ において, $\text{tf} \cdot \text{idf}$ を適用しない場合と $\text{tf} \cdot \text{idf}$ を適用する場合のいずれにおいても $\sin x$ を筆頭に \sin を含む数式が検索されている. これは, 変数の文字の違いを意識しなくても検索できることを示している.

問い合わせ $\max\{x, y\}$ において, $\text{tf} \cdot \text{idf}$ を適用する場合は, \max を含む数式が上位 5 位に 3 データ中 3 データ検索されているのに対し, $\text{tf} \cdot \text{idf}$ を適用しない場合では, \max を含む数式が上位 5 位に 2 データしか検索されていない. これは, $\text{tf} \cdot \text{idf}$ を適用することにより, どの数式にも多く含まれるノイズとなりうる可能性がある特徴についての重みが小さくなったため, より特徴が明確になったためであると考えられる.

次に問い合わせ $\sin t$, $\max\{x, y\}$, $\lim_{x \rightarrow a} x$, $\int_0^a (x + a)dx$ の場合の再現率 · 適合率のグラフをそれぞれ図 7, 8, 9, 10 に示す. これらからも, $\text{tf} \cdot \text{idf}$ を適用した場合の再現率 · 適合率が, $\text{tf} \cdot \text{idf}$ を適用しない場合よりも高くなっていることがわかる. また, $\text{tf} \cdot \text{idf}$ を適用する方式では, 全体的に再現率 · 適合率が高い.

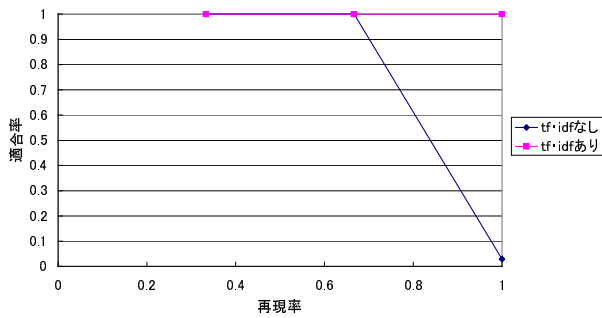


図 8 問い合わせ $\max\{x, y\}$ の場合の再現率・適合率.

Fig. 8 Recall rates and precision rates for the query $\max\{x, y\}$.

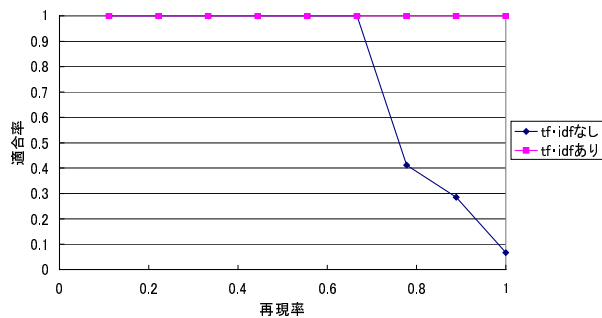


図 9 問い合わせ $\lim_{x \rightarrow a} x$ の場合の再現率・適合率.

Fig. 9 Recall rates and precision rates for the query $\lim_{x \rightarrow a} x$.

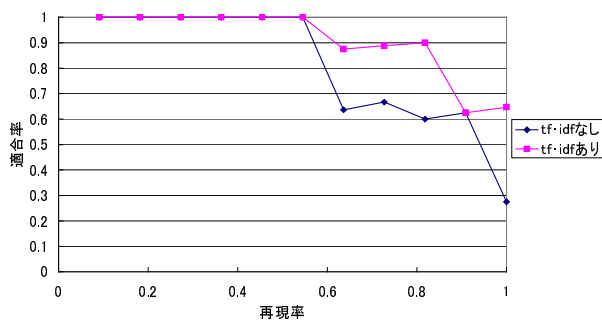


図 10 問い合わせ $\int_0^a (x+a)dx$ の場合の再現率・適合率.

Fig. 10 Recall rates and precision rates for the query $\int_0^a (x+a)dx$.

5.2 実験 2

5.2.1 実験環境

検索対象データとして, Maple を使って Content Markup で書き出した 23 個の数式, 計 400 個を用いる. Maple を使って書き出した式は, 数学公式・数表ハンドブック [10] から引用した.

検索対象データからのメタデータの抽出では, まず XML パーサを用いて DOM ツリーを生成した. この DOM ツリーからタグの出現頻度をカウントした. これらの処理では Java プログラムを用いた. また, 特異値分解には Matlab(version 5.2) [11] を用いた.

図 6 に実験システムの概要を示す. このシステムにおいてサーバー側の処理は Java Servlet により行った. 表 1 にはシステムに用いたアプリケーションを示す.

5.2.2 実験方法

5.2.3 実験結果

5.3 考察

数式からメタデータを抽出する際に $tf \cdot idf$ を適用することで, より明確な特徴を示すメタデータが抽出可能であることが分かった. また, 再現率・適合率は全体的に高い値となっており, 本類似数式検索の有効性を示している.

本方式では, 演算子間の関係については言及しておらず, 例えば \sin と \cos など比較的關係のある演算子についてもその他の演算子と同様に独立に扱われている. これについては, 意味の数学モデルによる意味的連想検索 [12], [13], [14] を本方式に適用することにより, \sin と \cos など演算子同士の意味的關係を反映した検索が可能になると考えられる. また, 本方式は演算子の出現順については考慮されていない. これらについては今後の課題である.

6. おわりに

本稿では, MathML で表された数式を対象とした類似数式検索の実現方式について示した. 本方式により, 数式の演算子に注目し, 添え字や変数に使う文字は違うが同様の意味を表す数式に関しても, 同じ公式として検索可能となった. さらに, 数式データからメタデータを抽出する際に, $tf \cdot idf$ を適用することで, より明確な特徴を示すメタデータが抽出可能となり, 問い合わせに合致した検索が可能となった. また, 実験システムを構築し実験を行い, 本方式の有効性を示した. 本方式により, これまでのドキュメントの内容による検索と本方式とを組み合わせることによる数式を含む科学論文などのドキュメント検索システムの実現が考えられ, ドキュメントの多角的な検索が期待できる.

今後の課題として, 本方式の意味の数学モデルによる意味的連想検索への適用, 数式のより詳細な構造を反映した意味的な計量方式の実現, および, 文章を対象とした検索システムとの統合方式の実現が挙げられる.

謝 辞

本研究を行うにあたり, 多くの助言をいただきました筑波大学理工学研究科の伊東拓氏に感謝いたします.

文 献

- [1] K.R. Foster, "Math on the Internet," IEEE Spectr., vol.36, no.4, April 1999.
- [2] "W3C Math Home," W3C.
<http://www.w3.org/Math/>
- [3] "MathML Software," W3C.
<http://www.w3.org/Math/implementations.html>
- [4] "TtM, a TeX to MathML translator," Ian Hutchinson.
<http://hutchinson.belmont.ma.us/tth/mml/>
- [5] M.W. Berry, and S.T. Dumais, and G.W. O'Brien, "Using linear algebra for intelligent information retrieval," SIAM Review, vol.37, no.4, pp. 573-595, December 1995.
- [6] P. Ion(editor), "MathML 2.0 implementation and interoperability report," Math WG Draft, February 2001.
<http://www.w3.org/Math/iandi/impl-interop20010220.html>
- [7] G. Salton, and C. Buckley, "Term-weighting approaches in automatic text retrieval," Inf. Process. and Management,

- vol.24, no.5, pp.513–523, 1988.
- [8] G. Salton, and C. Buckley, “Improving retrieval performance by relevance feedback,” *J. Am. Soc. Inf. Sci.*, vol.41, no.4, pp.288–297, June 1990.
 - [9] ”MathML Test Suite”, W3C.
<http://www.w3.org/Math/testsuite/>
 - [10] M.R. Spiegel, 数学公式・数表ハンドブック, 氏家勝巳 (訳), オーム社, 東京, 1995.
 - [11] ”The MathWorks: Developers of MATLAB and Simulink for Technical Computing”, The MathWorks, Inc.
<http://www.mathworks.com/>
 - [12] T. Kitagawa, and Y. Kiyoki, “The mathematical model of meaning and its application to multidatabase systems,” *Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems*, pp. 130–135, 1993.
 - [13] Y. Kiyoki, T. Kitagawa, and T. Hayama, “A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning,” *Multimedia Data Management – using metadata to integrate and apply digital media –*, McGrawHill, A. Sheth and W. Klas(editors), Chapter 7, 1998.
 - [14] 清木康, 金子昌史, 北川高嗣, “意味の数学モデルによる画像データベース探索方式とその学習機構,” *信学論 (D-II)*, vol.J79-D-II, no. 4, pp.509–519, April 1996.