

2R-8

数式処理システムGALにおける 数学公式データベース

佐々木建昭¹⁾, 三枝義典²⁾

1) 理化学研究所, 2) 図書館情報大学

1. はじめに

現在の数式処理ではアルゴリズムに基づく解法(決定的解法)が主流を占めているが, 決定的解法が無力である演算も数多い(例えば, 特殊関数を含む積分や無限級数など). 人間はこのような演算を発見的解法で解いているが, その鍵となるのは数学公式集の活用である. そこで我々は, 数学公式をデータベース化して数式処理システムに組み込み, システムで自動運用させることを考えた. こうすれば, 数式処理システムの機能の大幅な向上が期待できる.

本稿は, 上記の目的で開発した数学公式データベース(以下, 公式DBと略)の設計と構成, 及び機能について述べる. 公式DBは国産数式処理システムGAL(General Algebraic Language/Laboratory)上に開発されており, GALから自動運用すべく(GALのシステム記述言語で)インプリメントされている. なお, 現存のシステムではSMPのみが公式DBを装備しているが, その機能は限定されたものであって, 公式DBの自動運用の観点からは全く不十分である. その意味で, 本研究は公式DBの自動運用のさきがけとなるものである.

2. 数学公式データベースの設計

設計の基本方針として以下のようなものを設定した.

- (1) 公式データの更新は許さない. (このことは, 公式がユーザによって書き換えられる性格のものではないという認識に基づいている).
- (2) 公式は分野毎にまとめて格納し(これをサブデータベースと呼ぶ). 検索においては指定されたサブデータベースのみをサーチすることにより, メモリ使用効率と検索効率の向上を計る(このことは, 公式の実際の運用に際しても有効に作用すると考えられる).
- (3) 公式のインデックスは, 簡単でありながら実際の検索に便利なものを設定する(理論的完全さより実際的な有用性を重視する).
- (4) 検索文は人間が行なう公式集の手検索をなるべく忠

実に反映したものとする.

3. 公式のインデックス

上述のごとく, 我々の公式DBでは検索が主であり, しかも検索結果は余分な公式をなるべく含まないことが望ましい. そのため, インデックスの良し悪しが公式DBの出来を左右する. インデックス付けに際して, 我々は先ず公式を以下のように分類した(この分類に当てはまらない公式もあろうが, 数式処理システムで自動運用したい公式の大部分は以下の分類のいずれかに当てはまる).

【定義1: 公式の分類】第1種の公式とは, 積分や級数などのように, 決定的な演算に対して用いられる公式である. 第2種の公式とは, 三角関数の公式のように, 数式の簡単化を行なうために用いられる公式である. //

【定義2: キーワード】数式のキーワードとは数式中に含まれる“Integ”や“Sum”等の演算名や, “sin”や“exp”等の関数名とする. 数式中に関数名が含まれていなくて, “ π ”等の超越的数が含まれているならばこれをキーワードとする. //

【定義3: キーワードのレベル】 K_i を, 与えられた数式中の任意のキーワードとする. K_i が数式のトップレベルに出現しているならば, このキーワードのレベルを1とする. K_j が第 l レベルのキーワードであるとき, K_j の引数中にトップレベルで出現するキーワードのレベルを $l+1$ とする. //

各公式には, 以下の3つの型のインデックスを付与し(但しⅢ型のインデックスは名前付きの公式のみ), これらを併用して検索効率を上げる.

I型インデックススキーマ
= (公式番号, キーワード1, レベル)

II型インデックススキーマ
= (公式番号, キーワード2, 状態)

III型インデックススキーマ
= (公式番号, 名前)

①公式は, [左辺] \rightarrow [右辺] (条件) なる (条件付き) 書き換え規則と見なす.

Formula Database in Japanese Computer Algebra System GAL
Tateaki Sasaki¹⁾, Yoshinori Saigusa²⁾

1) Institute of Physical and Chemical Research,
2) University of Library and Information Science

② “キーワード1”とは【定義2】で定義したキーワードを指し，“キーワード2”とは【定義2】に全次数を意味する#DEG及び項数を意味する#TERMを付け加えたものである。

③ II型のインデックスは，公式の持つ“性質”を記述したものであり，“キーワード2”と“状態”との組み合わせで表現する。

④ 上述の“状態”とは，以下に述べるように第1種と第2種の公式では異なって定義される（これは実際の検索に都合良くするためである）。Rを任意の公式とすると，Rが第1種の公式であるならば[左辺]の主引数中に第1レベルで出現する任意のキーワード K_i ($i=1, \dots$)の次数を“状態”とする。例えば $R: \text{Integ}(\cos(X)) \rightarrow \sin(X)$ においては，RのII型インデックスは主引数“ $\cos(X)$ ”におけるキーワード（今の場合 \cos のみ）の次数で与えられる。Rが第2種の公式であるならば，Rの[左辺]と[右辺]のトップレベルに出現するキーワード K_i ($i=1, \dots$)の次数の増減関係(up, same, down, none)を“状態”とする。例えば $R: \sin(X+Y) \rightarrow \sin(X)\cos(Y) + \cos(X)\sin(Y)$ においては[左辺]と[右辺]における \sin と \cos の次数の増減関係がRのインデックスとなる。

なお，級数と積分（フーリエ変換等も含む）に対しては，級数のランニングインデックス（第n項のインデックスn）と積分変数もキーワードと見なして，より細かなインデックス付けを行なう。これは，例えば級数では，項が多項式や有理数で表わされる多数の公式があり，関数名のみをキーワードとしたのでは区別がつかぬからである。

4. 数学公式データベースの構成

公式DBは前節に述べた3種類のインデックステーブルと，公式自身を格納したテーブルから成る。公式テーブルのスキーマを以下に示す。

公式スキーマ = (公式番号, 左辺, 右辺, 条件)

これら4つのテーブルが各サブデータベース毎に作成され，まとめて管理される。但し，上で定義した3種類のインデックステーブルは，実際には検索効率を良くするために逆ファイル(inverted file)構造に再構成してデータベース内に格納されている。図1にI型とII型のインデックステーブルの例を示す。

5. 検索文

公式の検索には，システムが発するものとユーザが発する検索文の2種があるが，意味的にはどちらも同じである。現在は後者のみがインプリメントされており，前

者はGALのサポート機能が未完のためインプリメントするに至っていない。ユーザ用の検索文は以下の通り。

FIND<数式パターン>

WITH<パターン&検索条件>

ABOUT <固有名詞の列> ;

⑤ WITH節と ABOUT節は省略しても良く，

WITHはWHERE と書いても良い。

<数式パターン>とは，変数として<パターン変数>を含み得ることを除けば通常の数式そのものである。<パターン変数>とは任意の数式とマッチ可能な変数であり，GALでは“@”付きの変数として表わす。<パターン&検索条件>とは<パターン条件>and <検索条件>を意味し，それぞれ<数式パターン>に対する条件と検索したい公式の“性質”を記述する。<固有名詞の列>は，サブデータベース名や公式自身に付けられた名前等を指定する。図2に検索の例を示す。

6. おわりに

公式DBは単独ではその意義を見出すことはほとんど出来ないが，数式処理システム上で自動運用するとすればそのメリットは計り知れないものがある。この意味で，我々の研究はまだその一部を完了したにすぎない。今後は，公式運用モジュールを開発することにより数式処理システムで公式DBを実際に運用するための研究を進めなければならない。

公式番号	キーワード1	レベル
199	arcsin	1
200	arcsin	1
200	sin	2
200	sqrt	2

(図1a I型のインデックス)

公式番号	キーワード2	状態
120	#DEG	1
120	#TERM	1
120	exp	1

(図1b II型のインデックス-1種の公式)

公式番号	キーワード2	状態
97	#DEG	up
97	#TERM	down
97	cos	same
97	tan	none

(図1c II型のインデックス-2種の公式)

FIND SIN(@X)+SIN(@Y)

WITH UP(#DEG,COS) ABOUT TRIG;

(図2)

参考文献

- (1) 佐々木建昭, 増永良文, 三枝義典, 阿部昭博, 元吉文男, 佐々木睦子. “数式処理システムGALにおける数学公式データベース”, プログラミングシンポジウム予稿集, 1988.