ZAIXIANSHOUXIE

SHUXUE GONGSHI SHIBIE GUANJIAN JISHU YANJIU

在线手写数学公式识别关键技术研究

洪留荣 著

煤炭工业出版社

在线手写数学公式识别关键技术研究

洪留荣 著

煤炭工业出版社

·北 京·

图书在版编目（CIP）数据

在线手写数学公式识别关键技术研究/洪留荣著．—北京：煤炭工业出版社，2007.5

ISBN 978－7－5020－3005－6

Ⅰ．在… Ⅱ．洪… Ⅲ．数学-公式-机器识别-研究 Ⅳ．TP391.4

中国版本图书馆 CIP 数据核字（2007）第 004712 号

煤炭工业出版社 出版发行
（北京市朝阳区芍药居 35 号 100029）
网址：www.cciph.com.cn
北京房山宏伟印刷厂 印刷

*

开本 850mm×1168mm $^{1}/_{32}$ 印张 5 $^{1}/_{2}$
字数 139 千字 印数 1—1,000
2007 年 5 月第 1 版 2007 年 5 月第 1 次印刷
社内编号 5804 定价 18.00 元

版权所有 违者必究

本书如有缺页、倒页、脱页等质量问题，本社负责调换

内 容 提 要

本书针对在线手写数学公式识别，重点研究了数学公式字符识别、公式结构分析、数学公式标记语言等内容，并根据提出的数学公式逻辑结构模型，设计出一种简单易用的通用数学公式标记语言，用于公式的结构修改和基于文本的查询。

前　言

模式识别现已发展成为一门独立的新兴学科。目前国内外关于模式识别的著作不少，但还没有专门研究数学公式识别的书籍。

数学公式识别是在纸质数学文档转化为电子文档或创建电子科技文档时急需解决的一个现实问题；但数学公式是一种特殊识别对象，需要独特的理论和识别算法。

本书主要论述以下5个方面的内容：

一是应用笔划基元和空间关系的字符编码和模糊理论对字符进行同步分割和识别，以减少公式书写的约束。制定出一种规则，用以统一同一字符有不同笔顺时的编码，节省识别过程中的查询时间和识别样本的存储空间。

二是根据公式的结构特点，提出一种数学公式的逻辑结构模型，用于指导数学公式的逻辑结构识别和基于文本的公式查询。

三是在结构分析中应用字符凸壳和模糊理论判别字符间的空间关系，应用统计的方法重新划分其关系区域，给出了分割矩阵、行列式及条件公式等结构的理论和相关算法。

四是在公式整体结构识别中，对经典的最小生成树算法进行改进，在研究在线手写数学公式书写特征的基础上，提出新的权值计算方法，并对公式表结构进行了分析。

五是根据提出的数学公式逻辑结构模型，设计出一种简单易用的通用数学公式标记语言以指导公式结构修改以及实现基于文本的公式查询。

本书对提出的算法进行了详细描述，给出了实验的结果，并与其他算法结果进行了比较；详细地分析了以前一些流行算法的优点和不足之处，使读者更加全面地了解该领域的识别理论和方

此为试读，需要完整PDF请访问：www.ertongbook.com

法。

由于时间仓促，作者水平有限，书中难免有不足和失误之处，敬请专家、读者给予批评指正。

作　者

2007 年 5 月

目　录

1　绪论 …… 1

1.1　概述 …… 1
1.2　国内外研究现状 …… 5
1.3　公式识别系统存在的问题 …… 22

2　数学公式字符识别 …… 25

2.1　概述 …… 25
2.2　主要分割方法分析 …… 30
2.3　预处理 …… 35
2.4　笔划基元划分 …… 40
2.5　字符编码 …… 43
2.6　其他一些相关特征的提取 …… 54
2.7　字符数据搜集与训练 …… 56
2.8　分割和识别算法 …… 57
2.9　本章小结 …… 62

3　公式结构分析基础问题 …… 63

3.1　概述 …… 63
3.2　数学公式结构模型 …… 64
3.3　字符空间关系判别 …… 72
3.4　复合形分割 …… 88
3.5　本章小结 …… 104

4　公式结构分析 …… 106

4.1　概述 …… 106

4.2 支配关系 …… 107
4.3 数学公式的硬约规 …… 108
4.4 函数合并 …… 111
4.5 公式空间结构分析 …… 111
4.6 预处理后公式的结构分析 …… 112
4.7 实验结果与分析 …… 122
4.8 本章小结 …… 124

5 数学公式标记语言 …… 126

5.1 XML介绍 …… 126
5.2 数学公式标记语言 …… 129
5.3 XML关系存储 …… 137
5.4 相关方法 …… 140
5.5 中粒度关系型模型 …… 144
5.6 XML查询技术 …… 145
5.7 本章小结 …… 150

6 结论 …… 152

附录 …… 155
参考文献 …… 158

1 绪　论

1.1 概述

随着计算机技术的高速发展，人们可以快速处理大量的数据，如对大量信息进行存储、查询、统计等；但数据的输入速度远远低于数据的处理速度，这在很大程度上降低了计算机的使用效率。为了解决数据的输入问题，人们做了长期的积极的努力，提出了许多方法[1,2]。其中，基于键盘的输入方法是实现输入的主要手段；但人们在拟定文档时，为了集中注意力完成文档，往往更喜欢用笔和纸书写而不采用键盘。几十年来，许多研究者提出了大量手写输入的研究成果，也开发出了针对脱机识别印刷体字符、手写体字符和在线手写字符等许多识别系统。虽然在有些字符识别领域并不尽如人意，如脱机自由手写汉字，但仍有许多研究者在这些领域进行积极的研究和探索，以适应不同情况的要求[3~5]。也有一些很优秀的识别系统达到了很高的水平并在市面上流行，如汉王在线手写识别系统[6]。

对自然科学工作者特别是数学工作者来说，他们更关心数学公式的手写输入问题，因为这些系统并不能解决数学公式输入问题，他们中的不少人仍然使用键盘鼠标方式输入公式。

Paolo Bottoni 和 Bernd Meyer 等人把可视语言看成是一组图[7]，一个已定义语言的有效表达，可视语言的语法和语义通过图中字符间的空间关系和逻辑关系确定。然而，一些可视语言没有正规的语法定义，在应用时，某些符号的意义可以由使用者自己定义，这种定义可以被别人接受。数学公式正是这样的一种自然语言，因此数学公式可以看成是一种自然可视语言。

数学公式是用字符和符号的二维结构来表达有关数学思想

的，这种形式有助于数学思维，也方便观看。同样的，人们用这种结构与计算机通信也是自然和方便的，这种方便包括数学公式与内部计算机表达之间的通信。在当前的技术条件下，计算机很容易产生二维结构的数学公式；但由于识别技术还不是很成熟，把数学公式转换成计算机能处理的形式常常要靠用户参与处理。在工程科技领域，相对于串语言，用数学公式与计算机通信可以更加简单和有效地表达用户的意图，如用户用 Latex[8] 或 MathML[9] 这样的串语言表达一个数学公式比直接用数学公式表达要困难得多。这方面的例子有很多，如工程图等。

所以，一个好的数学公式识别系统可以提高计算机在公式中的处理能力，也可减轻人们输入公式的负担。目前，存在着两种把数学公式与计算机能处理的形式之间的转换方式，第一种是人工转换，第二种是自动转换，如图 1—1 所示。

在人工转换方式中，鼠标与键盘作为重要的输入工具，使用了很多年；但它们作为把数学公式输入计算机的主要手段还存在很多问题，这主要是因为自然语言中存在大量的字符，使字符输入变得很复杂。就数学公式而言，仅常用的字符就有 150 个以上[10]，如果考虑出现在一些专业数学中的字符和不同语系的人使用的字符，数量会更大。即使键盘和鼠标能解决这些字符的输入，其速度也非常慢。目前能输入数学表达式的成熟系统全部采用键盘鼠标输入；同时配以相应的软件，有的需要长时间地学习、记忆大量的控制命令，如 Latex[8]；有的由于编辑软件（如 Word）需要频繁切换到插入符号方式，所以输入速度很慢。在自动转换方式中，对在线识别和离线识别的研究较多，基于语音识别的研究较少。

一般来讲，识别系统所要达到的要求主要看它所提供的应用，针对不同的应用系统，所用到的方法和原理差别较大。一般来说，数学公式识别系统主要有两个作用。

一是用于个人创建要打印的文档。数学公式在线识别系统和语音识别系统主要应用于这个方面。在线识别系统应用手写板和

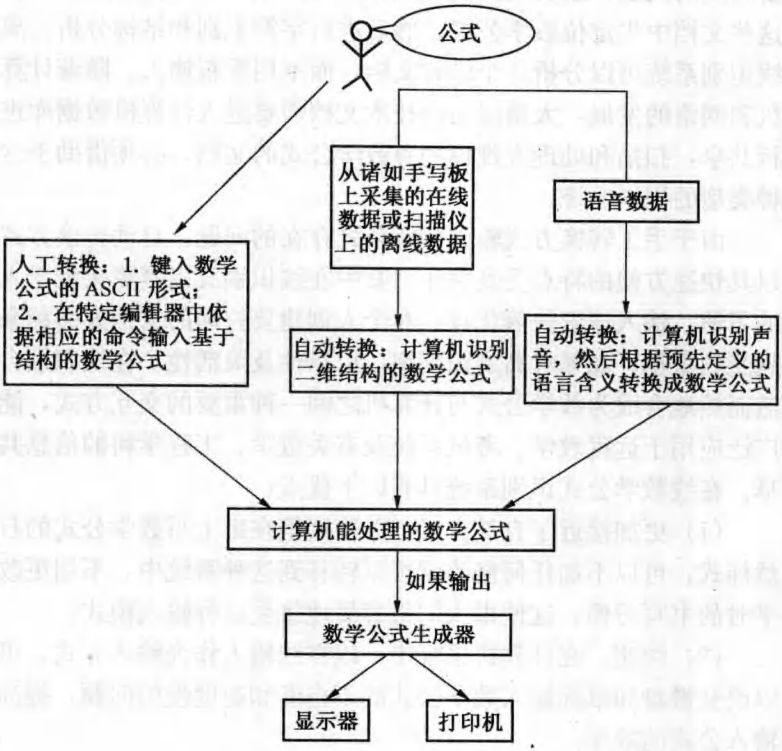

图 1—1　数学公式与计算机能处理的形式之间的转换方式

电子笔输入公式，它要求识别器能处理数学公式字符和分析公式的结构。这对个人输入要打印的文档是一种有效的手段，这种系统具有自然、快速的特点。语音识别系统通过麦克风、声卡等语音输入设备把声音送入计算机，再由计算机根据所给定的语法生成公式，这种方式的一个很大的好处是适用于书写有困难的人群。由于数学公式是一种自然语言，这就决定了研究者必须构建说公式的语法和规则，用户也要掌握这些语法和规则。

二是用于纸质文档的数字化。这是数学公式离线识别系统主

此为试读，需要完整PDF请访问：www.ertongbook.com

要解决的问题，系统先扫描以前的文档（打印的或手写的），在这些文档中先定位数学公式，然后进行字符识别和结构分析。离线识别系统可以分析一个已有文档，而不用重新输入。随着计算机和网络的发展，大量以前的技术文档需要进入计算机数据库进行共享，扫描和处理大规模的有数学公式的文档，必须借助于这种类型的识别系统。

由于手工转换方式输入数学公式存在的问题，自动转换方式以其快速方便的特点受到关注。其中在线识别又以其更接近于人的习惯、输入速度快等优点，在个人创建要打印的文档方面越来越受到重视。如果有高的识别率、可靠性及灵活性，在线识别自然而然地会成为数学公式与计算机之间一种重要的交互方式，能广泛应用于远程教学、考试系统及有关数学、工程学科的信息共享。在线数学公式识别系统具有以下优点：

（1）更加接近于自然方式。以往用笔在纸上写数学公式的自然样式，可以不加任何修改就可以转移到这种系统中，不用更改平时的书写习惯。这使得人们很容易地接受这种输入模式。

（2）快速。在计算机系统中，以在线输入作为输入方式，可以改变键盘和鼠标输入数学公式的不自然和速度慢的问题，提高输入公式的效率。

随着互联网的普及、地域概念的淡化、多媒体技术的进步，近几年来网络信息共享需求增加，在普遍使用的 HTML（超文本标记语言）网页模式中，要实现理工科中大量使用的数学符号界面还是比较困难的。现在一般以图片的方式显示或者以 PDF 文件格式显示，这样的方式难以对数学公式进行修改、查询和重新利用，从而限制了对资源的使用。这一问题在现代远程教育模式的讨论中已是数不胜数。虽然 Latex 和 MathML 等可以实现网上公式的查询，但它们着重于描述公式的显示细节，这也就决定了它们在公式查询和重新利用方面的不足。

综上所述，研究数学公式识别系统可以减轻人们输入数学公式的负担，增加输入的速度，使输入方式更加自然，同时，可以

使数学公式的信息交流更加方便。

1.2 国内外研究现状

国内外研究数学公式的输入已有很长时间，本节首先对已有数学公式的输入模式作一个简单的概述，然后介绍在线数学公式输入系统的国内外研究现状，给出以前此领域研究的主要工作，对其中一些工作的原理和方法作适当比较。

1.2.1 数学公式输入系统的几种模式

数学公式输入系统以其输入形式可以分成模板输入、键盘命令输入、排版输入、手写识别输入和语音输入 5 种模式[11]。

(1) 模板输入系统。这种输入方式是用户通过与计算机之间交互，选取计算机事先提供的数学公式结构模板，然后依据相应的结构输入公式。程序一般提供一个窗体，窗体放入模板和一些命令按钮，每一个命令按钮表示数学公式的一个基本结构或字符，基本结构让用户输入其他字符或填充新的基本结构。这种模式有两个优点：其一，这种强制性的输入可以保证公式结构最优；其二，只需通过简单加入公式和字符模板就可以扩充现有的输入功能。但这种模式也有其不利的一面，多模板的输入会使用户不断地在键盘和鼠标之间切换，所以输入速度很慢，典型的有 Word 中数学公式编辑器。

(2) 键盘命令输入系统。一些公式输入系统如 Emacs[12]，可以让用户完全使用键盘输入数学公式。由于一个键盘上的字符有限，不能表达所有的数学公式字符，所以这类系统设计了专门的基于键盘的字符输入命令，这就使得用户在应用这种系统之前必须学习相应的键盘命令，而且这种系统要求用户敲击字符命令和敲击公式基本结构时要进行切换，所以输入速度同模板输入速度一样也很慢，且用户要花相当多的时间来学习有关操作指令。

(3) 排版输入系统。排版输入系统如 Latex，虽然它可用文本描述数学公式，具有很好的结构，可以处理复杂的数学公式，但系统的语法复杂，一般要花很长时间才能掌握，而且修改困

难。目前一般都是在相应的应用软件上用鼠标和键盘进行公式编辑，而很少有人直接应用其文本方式输入公式。

(4) 手写识别输入系统。这种系统为用户提供了一个自然的没有约束的数学公式输入方式。系统通过电子笔输入公式，计算机采集输入信息，通过识别系统转化成计算机可以理解的公式数据和模式，因此受到越来越多的关注。其主要优点是不需要用户长时间地学习，输入自然且速度快。当前，虽然有一些系统被开发出来，但离实际应用还有较大差距。

(5) 语音输入系统。这是一种特别的输入方式，它首先定义怎样说数学公式，即定义说数学公式的语法，然后通过语音输入数学公式，计算机识别语音后，依据所定义的语法还原数学公式。这种输入方式不仅适用于不能使用鼠标和键盘的用户，也适用于一般用户，但这种方式的缺点在于需要输入者掌握所定义的说数学公式的语言。例如，Aster[12]系统应用 Latex 作为语言以语音方式输入数学公式。由于 Latex 语法结构复杂，系统没有得到广泛认可。

1.2.2 手写数学公式的识别过程

数学公式以字符和符号的二维方式排列来传达它的数学含义，正如一个人看到一个公式一样，首先是认识公式中的字符和符号，然后再对其结构加以分析，以确定一个公式所要表达的意义。有时也可以用后面的结构来理解前面的一些模糊的字符或结构。计算机识别的过程也遵循这一规则，在字符识别时，把输入的数据确定为单一的字符和符号，在结构分析时处理这一组字符和符号的二维空间关系以显示所给定的数学公式的信息。所以数学公式的识别涉及两个方面的内容：一是字符识别，一是结构分析。而字符识别又与一般的文本识别不同，数学公式中的字符排列是二维的，而一般文本的识别是一维的，这显然增加了数学公式中字符识别的难度。

尽管存在很多二维识别系统，在原理上可移植到数学公式识别中，但数学公式的特殊性决定了它们必须经过本质的修改[13]。

虽然数学公式的识别研究已经有很多年[14]，但只是到近几年才有更多的研究者关注这一领域。这些识别系统均包括字符识别和结构分析这两个步骤，Blostein[15]又把这两个步骤进一步细分为以下 6 个小的步骤：

（1）预处理：去除笔划噪声和笔划扭曲等。

（2）字符分割：把公式中的各个字符独立出来，即确定哪些笔划数据属于同一个字符。

（3）识别字符：针对分割的数据识别出数学公式的字符和符号。

（4）判别字符之间的空间关系：数学公式中字符和符号之间的关系从空间结构来看，可分为正上、正下、左上、左下、右上、右下、左水平、右水平和内部等关系。

（5）判别字符之间的逻辑关系：字符之间从逻辑关系上看存在着关系，有些字符之间的关系从空间上看可能是一种关系，但从逻辑上看是另一种关系。如 $\sum_{i=2}^{10000} i$ 中“1”与“Σ”，从空间来看是一种左上关系，而从逻辑关系来看是一种正上关系。

（6）分析公式的语义：分析数学公式中符号与结构的含义，前面提到数学公式是一种自然语言，所以对所有公式定义语义是很困难的事情。

这 6 个步骤的前 3 个步骤完成字符识别，后 3 个步骤完成结构分析。在具体的识别过程中这 6 个步骤可以顺序执行，也可以交叉执行，用后一个阶段所得出的信息为前一个阶段提供服务，所以在识别顺序上没有一个明显的界限，也可以先完成局部的空间关系识别，然后再进行后续的字符分割和识别[16]。

在研究中，专门集中于某一个或几个步骤而忽略其他步骤是有可能出现的。例如，可以专门针对数学公式的结构加以研究，而认为前面的字符识别阶段完全正确[13]。在实际应用的识别系统中还不可能达到字符识别完全正确，所以在结构分析之前通常以人工方式对字符的识别错误加以更正，以保障后续结构分析的

效果。一个优秀的识别器，应该具有以下几个特点：

（1）有自动的字符分割和识别功能且性能较强。自动分割识别字符是整个数学公式识别系统的基础，字符分割和识别的错误对后续工作影响非常大。它造成结构分析的困难，并会造成结构分析的错误，如果用户过多地介入字符纠正，会在很大程度上影响使用速度。

（2）有修改公式结构结果的功能。数学公式中字符间的关系相当复杂，二义性也较多，这就要求系统有好的结构分析方法且能提供方便的修改错误的功能。

（3）具有针对各种公式的结构分析功能。一个识别器如果只是针对公式的某一子集有好的性能，并不能认为这是一个好的识别器。

1.2.3 在线数学公式输入系统的研究现状

1.2.3.1 字符分割

分割是对字符数据进行分解、分组，使之成为一个独立的可分类数据单元，即这个单元可表示为一个单一的字符。对于在线识别系统，字符的分割就是把手写板和电子笔输入的笔划分组为单个的字符。由于数学公式字符存在二维关系且包括不同语系的字符和专门的数学公式符号，使分割成为一个非常复杂的问题。

典型的字符分割方法就是把物理上分开的部件归为不同的组，但在数学公式中，许多字符是由多个部件组成的，如“i”、“j”、“=”等，所以在识别字符之前，要把这些相关的部件组织在一起。更复杂的情况是，如根号这样的字符其本身就包含着其他字符，这也增加了字符分割的困难程度。

以下几种方法用于分割很具代表性。Faure 和 Wang[16] 用基于数据驱动的分割模块和基于知识驱动的分割模块对字符进行分割。基于数据驱动的分割模块用于建立给定数学公式的结构关系树，应用 x 轴和 y 轴的投影来分割数据，由于投影方法不适合于像分式和根号这样的符号，所以另外应用了一个移动操作算法。在投影之前，先对这些特殊符号进行处理，把这些特殊符号

抽取出来，再利用 x 和 y 轴的投影来分割字符，以形成一个字符的关系树。基于知识驱动的分割模块主要用于前一个模块生成的树结构，由于数学公式中融合了一些特定字符如“=”、“i”、“j”等，它们在物理上是断开的，所以在这个模块中调用它们在前一个模块中形成的关系树，把这些笔划分为一组。

Ha[17]等应用“x—y 递归”的分割方法进行印刷体公式的字符分割，x 指水平方向分割，y 指垂直方向分割。这种方法利用的是字符的边界框在 x 轴和 y 轴上的递归投影来分割字符，Okamoto[18]等也通过水平和垂直投影的分割方法进行分割，但它们是基于像素的。这两种方法在原理上都可用于在线数学公式识别。

Smithies[19]用一种简单渐进分组算法进行在线字符分割，这个系统首先为笔划进行尽可能的分组，然后，由字符识别器事先给定的最大可信度值来决定分组形式。这种算法简单、快速，但经常会引入错误，需要人工修改。

ChamFai Chen 应用窗体方法进行字符的分割[20]。Winkler[21]等从给定的笔划集中创建一个字符假设网（SHN：Symbol Hypothesis Net），在字符分组时，SHN 得出各种笔划组合的可能性，在 SHN 中每一种笔划组合代表一种可能的字符，对这些组合用 HMM（Hidden Markov Model）进行分类，分类器的概率结果与每一种笔划组合似然值相关联，最大似然值对应的结果作为最后的分割结果。

Toyozumi[22]等提出了基于笔划距离和公式结构信息的字符分割方法，他们在动态规划方法中应用了一个纠正值和一个评估值，纠正值由笔划之间的位置和公式的结构计算。Xiangyang Feng 应用边界框决定的方法[23]分割字符。

这些算法在结构不太复杂且书写适应一定规则的数学公式中，进行小字符集的分割有较好的效果；但存在的问题明显，主要是这些算法所限制的规则和范围在实际中很容易被突破，造成这些算法并不适合于实际应用。

此为试读，需要完整PDF请访问：www.ertongbook.com