



中国计算机学会  
学术著作丛书

# 语言信息处理专论

黄昌宁 夏莹 主编



清华大学出版社  
广西科学技术出版社

**(京)新登字 158 号**

## **内 容 简 介**

本书收入的 20 篇论文选自清华大学信息科学技术学院计算机、电子工程和自动化等三系的研究生学位论文。全书按内容分成汉语文本的分析与生成、汉字识别和汉语语音识别与合成等三篇。书中重点介绍了清华大学在此领域中取得的科研成果及研究方法。

第一篇讨论了汉语自动分词、词性标注和词义辨识,介绍了汉语句法分析系统不同的实现方法,汉语句子和话语(句群)的生成技术。

第二篇讨论了汉字识别的特点、类型和方法,并介绍了基于“人工神经网络”的汉字识别方法。

第三篇讨论了特征抽取、模式划分、时间对准等问题,建立了识别用的标准模型,并对神经网络用于汉语语音识别进行了研究。

本书对从事语言信息处理的研究人员是一本难得的好参考书。

**版权所有,翻印必究。**

**本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。**

## **图书在版编目(CIP)数据**

语言信息处理专论/黄昌宁,夏莹主编。—北京:清华大学出版社,1995  
ISBN 7-302-01929-0

I. 语… II. ①黄… ②夏… III. 语言信息处理学-汉字信息处理 IV. TP391

中国版本图书馆 CIP 数据核字(95)第 11796 号

出版者:清华大学出版社(北京清华大学校内,邮编 100084)

广西科学技术出版社(南宁市河堤路 14 号,邮编 530021)

印刷者:北京人民文学印刷厂

发行者:新华书店总店北京科技发行所

开 本:787×1092 1/16 印张:41.75 字数:987 千字

版 次:1996 年 4 月第 1 版 1996 年 4 月第 1 次印刷

书 号:ISBN 7-302-01929-0/TP·879

印 数:0001—1000

定 价:56.00 元

清华大学出版社 广西科学技术出版社  
计算机学术著作出版基金

## 评 审 委 员 会

主 任 委 员 张效祥

副主任委员 周远清 汪成为

委 员 王鼎兴 杨芙清 李三立 施伯乐 徐家福

夏培肃 董韫美 张兴强 徐培忠

## 出版说明

近年来,随着微电子和计算机技术渗透到各个技术领域,人类正在步入一个技术迅猛发展的新时期。这个新时期的主要标志是计算机和信息处理的广泛应用。计算机在改造传统产业,实现管理自动化,促进新兴产业的发展等方面都起着重要作用,它在现代化建设中的战略地位愈来愈明显。计算机科学与其它学科的交叉又产生了许多新学科,推动着科学技术向更广阔的领域发展,正在对人类社会产生深远的影响。

科学技术是第一生产力。计算机科学技术是我国高科技领域的一个重要方面。为了推动我国计算机科学及产业的发展,促进学术交流,使科研成果尽快转化为生产力,清华大学出版社与广西科学技术出版社联合设立了“计算机学术著作基金”,旨在支持和鼓励科技人员,撰写高水平的学术著作,以反映和推广我国在这一领域的最新成果。

计算机学术著作出版基金资助出版的著作范围包括:有重要理论价值或重要应用价值的学术专著;计算机学科前沿探索的论著;推动计算机技术及产业发展的专著;与计算机有关的交叉学科的论著;有较大应用价值的工具书;世界名著的优秀翻译作品。凡经作者本人申请,计算机学术著作出版基金评审委员会评审通过的著作,将由该基金资助出版,出版社将努力做好出版工作。

基金还支持两社列选的国家高科技重点图书和国家教委重点图书规划中计算机学科领域的学术著作的出版。为了做好选题工作,出版社特邀请中国计算机学会、中国中文信息学会帮助做好组织有关学术著作丛书的列选工作。

热诚希望得到广大计算机界同仁的支持和帮助。

清华大学出版社  
广西科学技术出版社 计算机学术著作出版基金办公室

1992年4月

## 序 言

计算机是当代发展最为迅猛的科学技术,其应用几乎已深入到人类社会活动和生活的一切领域,大大提高了社会生产力,引起了经济结构,社会结构和生活方式的深刻变化和变革,是最为活跃的生产力之一。计算机本身在国际范围内已成为年产值达 2500 亿美元的巨大产业,国际竞争异常剧烈,预计到本世纪末将发展为世界第一大产业。计算机科技具有极大的综合性质,与众多科学技术相交叉而反过来又渗入更多的科学技术,促进它们的发展。计算机科技内容十分丰富,学科分支生长尤为迅速,日新月异,层出不穷。因此在我国计算机科技尚比较落后的情况下,加强计算机科技的传播实为当务之急。

中国计算机学会一直把出版图书刊物作为学术活动的重要内容之一。我国计算机专家学者通过科学实践,做出了大量成果,积累了丰富经验与学识。他们有撰写著作的很大积极性,但相当时期以来计算机学术著作由于印数不多,出版往往遇到不少困难,专业性越强越有深度的著作,出版难度越大。最近清华大学出版社与广西科学技术出版社为促进我国计算机科学技术及产业的发展,推动计算机科技著作的出版工作,特设立“计算机学术著作出版基金”,以支持我国计算机科技工作者撰写高水平的学术著作,并将资助出版的著作列为中国计算机学会的学术著作丛书。我们十分重视这件事,并已把它列为学会本届理事会的工作要点之一。我们希望这一系列丛书能对传播学术成果、交流学术思想、促进科技转化为生产力起到良好作用,能对我国计算机科技发展具有有益的导向意义,也希望我国广大学会会员和计算机科技工作者,包括海外工作和学习的神州学人们能积极投稿,出好这一系列丛书。

中国计算机学会

1992 年 4 月 20 日

## 前 言

在我国计算机与信息产业中,汉语汉字处理技术具有重要地位。拥有良好的汉语汉字处理能力的计算机产品不仅具有广阔的国内市场,而且总是优先被世界大市场所接受。

语言文字是人类信息和知识的主要载体,因此语言信息处理的技术水平已成为衡量一个国家现代化、信息化水平的重要标志。由于汉语汉字本身的特异性质,汉语汉字处理技术包含着许多不同于印欧语系语言文字处理技术的特点。这就是我们编辑这部专论的初衷。

本书收入的 20 篇论文选自清华大学信息科学技术学院计算机、电子工程和自动化等三系的研究生学位论文。全书按内容分成汉语文本的分析与生成、汉字识别和汉语语音识别与合成等三篇。黄昌宁、夏莹和方棣堂分别负责为各篇挑选论文和撰写综述。

“严谨,勤奋,求实,创新”历来是清华的校训。清华大学研究生的学位论文除了其研究方法和成果以外,是否还能给读者留下一些可资借鉴的别的什么印象吗?

黄昌宁 夏 莹

1994 年 12 月于清华园

## PREFACE

The Chinese character and language processing technology plays an important role in our computer and information industries. Computer products with well-designed Chinese character and language processing capacity have not only a broad market in our country, but also a preference of acceptance in the worldwide market.

Languages and writings are the main carrier of human information and knowledge, consequently the technical level of linguistic information processing has become an important sign for measuring the modernization and informatary level of a country. As the idiosyncrasy of the Chinese character and language themselves, the Chinese character and language processing technology involves many features different from those of Indo-European writings and languages.

There are twenty papers in the book, which are all graduates' theses selected from Departments of Computer Science, Electronic Engineering, and Automation, the Institute of Information Science and Technology, Tsinghua University. According to their contents, those theses can be divided into three parts: (1) Chinese text analysis and generation; (2) Chinese character recognition; and (3) Chinese speech recognition and synthesis. Professors Huang Changning, Xia Ying and Fang Ditang are responsible respectively for selecting papers and making surveys for each part of the book.

"Rigorous, realistic and creative" are the admonition of Tsinghua for a long time. Except their research methods, whether those theses of Tsinghua University can leave some impressions on readers that they can make use of?

Huang Changning and Xia Ying  
Tsinghua Yuan  
December 1994

# 目 录

## 第 1 篇 汉语文本的分析与生成

|                          |         |
|--------------------------|---------|
| 基于评价的汉语自动分词系统的研究与实现..... | 马 晏 (2) |
| 第 1 章 引言.....            | (3)     |
| 1.1 问题的提出.....           | (3)     |
| 1.2 书面汉语自动分词的目标.....     | (3)     |
| 1.3 历史回顾.....            | (4)     |
| 1.4 目前存在的问题.....         | (5)     |
| 1.5 研究的目标.....           | (5)     |
| 第 2 章 汉语自动分词的形式模型.....   | (6)     |
| 2.1 切分的形式模型.....         | (6)     |
| 2.2 切分歧义的类型.....         | (8)     |
| 2.3 最大匹配法.....           | (9)     |
| 2.4 全切分 .....            | (10)    |
| 2.5 评价问题 .....           | (11)    |
| 2.6 本章小结 .....           | (11)    |
| 第 3 章 汉语自动分词的评价层次 .....  | (12)    |
| 3.1 评价的提出 .....          | (12)    |
| 3.2 词典的作用 .....          | (12)    |
| 3.3 经验和统计的意义 .....       | (13)    |
| 3.4 句法和语义一级的评价 .....     | (13)    |
| 3.5 语境一级的评价 .....        | (14)    |
| 3.6 本章小结 .....           | (14)    |
| 第 4 章 词典的完备性与完全性 .....   | (14)    |
| 4.1 问题的提出 .....          | (14)    |
| 4.2 词典的完备性 .....         | (14)    |
| 4.3 词典的完全性 .....         | (15)    |
| 4.4 词典与分词标准的一致性 .....    | (19)    |
| 4.5 本章小结 .....           | (19)    |
| 第 5 章 自动分词策略 .....       | (20)    |

|       |                         |      |
|-------|-------------------------|------|
| 5.1   | 主要算法回顾及切分盲点 .....       | (20) |
| 5.2   | 全切分算法 .....             | (21) |
| 5.3   | 修剪问题 .....              | (23) |
| 5.4   | 本章小结 .....              | (26) |
| 第 6 章 | 几种评价参数 .....            | (26) |
| 6.1   | 评价参数的意义 .....           | (26) |
| 6.2   | 评价参数 .....              | (27) |
| 6.3   | 本章小结 .....              | (29) |
| 第 7 章 | 一个基于评价的自动分词系统 SEG ..... | (30) |
| 7.1   | 系统组成 .....              | (30) |
| 7.2   | 词典的构造 .....             | (30) |
| 7.3   | SEG 的分词策略 .....         | (32) |
| 7.4   | 切分效果分析 .....            | (33) |
| 7.5   | 本章小结 .....              | (34) |
| 第 8 章 | 结束语 .....               | (34) |
| 8.1   | 切分正确率问题 .....           | (34) |
| 8.2   | 自动分词系统的作用与发展 .....      | (35) |
| 8.3   | 计算语言学的前景 .....          | (35) |
|       | 参考文献 .....              | (36) |

## **基于统计的汉语语料库词性自动标注的研究与实现 .....** 白栓虎 (37)

|       |                           |      |
|-------|---------------------------|------|
| 第 1 章 | 绪论 .....                  | (38) |
| 1.1   | 计算语言学的研究方法 .....          | (38) |
| 1.2   | 语料库和基于语料库的方法 .....        | (39) |
| 1.3   | 语料库词性标注及其意义 .....         | (41) |
| 1.4   | 问题的描述 .....               | (41) |
| 1.5   | 词性标注的研究方法及现状 .....        | (43) |
| 1.6   | 本论文主要工作内容 .....           | (44) |
| 第 2 章 | 语料库语法标记集的设计 .....         | (44) |
| 2.1   | 制定标记集的标准和原则 .....         | (44) |
| 2.2   | 汉语语料库词性标注标记集 .....        | (45) |
| 2.3   | 对标记集所作的修改 .....           | (48) |
| 2.4   | 本章小结 .....                | (49) |
| 第 3 章 | 统计语言学模型 .....             | (49) |
| 3.1   | 理论依据 .....                | (50) |
| 3.2   | n 元语法(n-gram)及其实现方法 ..... | (51) |
| 3.3   | 数据稀疏与插值马尔科夫模型 .....       | (52) |
| 3.4   | 经验统计语言学模型 .....           | (53) |

|      |                                   |      |
|------|-----------------------------------|------|
| 3.5  | 本章小结 .....                        | (53) |
| 第4章  | 标记选择 .....                        | (53) |
| 4.1  | 词性标注的动态规划算法 .....                 | (53) |
| 4.2  | 修正的 Forward-Backward 算法 .....     | (55) |
| 4.3  | 对 FB 法输出的讨论 .....                 | (56) |
| 4.4  | 输出实例 .....                        | (57) |
| 4.5  | 本章小结 .....                        | (58) |
| 第5章  | 影响正确率的因素及词典空缺的处理 .....            | (59) |
| 5.1  | 标注试验的方法 .....                     | (59) |
| 5.2  | 训练集的规模与同现标记对、词典的关系 .....          | (59) |
| 5.3  | 模型中不同因素的地位 .....                  | (61) |
| 5.4  | 词典的作用和词典空缺 .....                  | (63) |
| 5.5  | 词典空缺处理策略 .....                    | (63) |
| 5.6  | 词典应具备的规模 .....                    | (65) |
| 5.7  | 推测实例 .....                        | (65) |
| 5.8  | 本章小结 .....                        | (66) |
| 第6章  | 系统的实现与测试 .....                    | (66) |
| 6.1  | 标记集和语料 .....                      | (67) |
| 6.2  | MMCRT 系统的组成 .....                 | (67) |
| 6.3  | 标注系统的词典 .....                     | (67) |
| 6.4  | 模型训练 .....                        | (69) |
| 6.5  | 标注模块 .....                        | (70) |
| 6.6  | 语料库管理 .....                       | (71) |
| 6.7  | 各种模型的测试 .....                     | (71) |
| 附录1  | “八五”汉语语料库的词性标记集(1990年6月第二稿) ..... | (72) |
| 附录2  | 对第2章细类合并得到的大类(TAGSET2) .....      | (75) |
| 参考文献 | .....                             | (76) |

## 汉语真实文本的语义自动标注 ..... 童翔 (78)

|     |                     |      |
|-----|---------------------|------|
| 第1章 | 引言 .....            | (79) |
| 1.1 | 词语歧义及歧义排除 .....     | (79) |
| 1.2 | 语料库和语料库的标注 .....    | (83) |
| 1.3 | 汉语的义项标注 .....       | (84) |
| 第2章 | 汉语语义标注系统 .....      | (86) |
| 2.1 | 资源 .....            | (86) |
| 2.2 | 系统的组成和主要数据结构 .....  | (89) |
| 2.3 | 系统工作流程 .....        | (90) |
| 第3章 | 基于例子的汉语义项标注方法 ..... | (91) |

|       |                            |       |
|-------|----------------------------|-------|
| 3.1   | MTD 与实例库 .....             | (91)  |
| 3.2   | 基于实例的汉语义项标注算法 .....        | (92)  |
| 第 4 章 | 系统的评价与展望 .....             | (93)  |
| 4.1   | 系统的评价 .....                | (93)  |
| 4.2   | 系统的局限性及改进 .....            | (94)  |
| 4.3   | 今后工作的展望 .....              | (96)  |
| 结论    | .....                      | (97)  |
| 附录 1  | 标注样本 .....                 | (98)  |
| 附录 2  | 《同义林词林》大类和中类的代码和分类名称 ..... | (100) |
| 参考文献  | .....                      | (101) |

## 汉语句法分析中的一种多扫描确定性算法及其在篇章理解中的应用..... 孙茂松 (103)

|       |                          |       |
|-------|--------------------------|-------|
| 第 1 章 | 绪论.....                  | (104) |
| 1.1   | 句法分析概述.....              | (104) |
| 1.2   | 篇章理解概述.....              | (107) |
| 1.3   | 句法分析与篇章理解的关系.....        | (108) |
| 1.4   | 本文的任务及内容.....            | (109) |
| 第 2 章 | 汉语中的兼类词、同形类组及其处理策略 ..... | (109) |
| 2.1   | 兼类词、同形词组及同形类组 .....      | (109) |
| 2.2   | 兼类词及其处理策略.....           | (112) |
| 2.3   | 同形类组及其处理策略.....          | (115) |
| 第 3 章 | 汉语句法分析的困难.....           | (120) |
| 3.1   | 困难之一——汉语句法结构之间界限模糊.....  | (120) |
| 3.2   | 困难之二——兼类词和同形类组.....      | (122) |
| 3.3   | 困难之三——差异的普遍性.....        | (122) |
| 3.4   | 困难之四——语义的引入.....         | (124) |
| 3.5   | 走出困境的几点措施.....           | (124) |
| 第 4 章 | 算法构思.....                | (127) |
| 4.1   | 算法宜采用确定性.....            | (129) |
| 4.2   | 控制策略:自底向上与数据驱动 .....     | (129) |
| 4.3   | 从左向右、从右向左多次扫描 .....      | (130) |
| 4.4   | 两个“贯穿”.....              | (137) |
| 第 5 章 | 算法及实验系统.....             | (138) |
| 5.1   | 数据结构.....                | (138) |
| 5.2   | 汉语句法分析的一种多扫描确定性算法.....   | (141) |
| 5.3   | 实验系统.....                | (144) |
| 第 6 章 | 典型军事情报的篇章理解.....         | (145) |
| 6.1   | 一类典型军事情报的特点.....         | (145) |

|     |                     |       |
|-----|---------------------|-------|
| 6.2 | 军事情报理解中的几个问题        | (146) |
| 6.3 | 程序实现                | (149) |
| 第7章 | 结束语                 | (151) |
| 附录1 | 句法分析部分结果(单句)        | (151) |
| 附录2 | 句法分析部分结果(军事情报篇章)    | (160) |
| 附录3 | 一篇军事情报的内部表示(机器自动抽取) | (163) |
|     | 参考文献                | (168) |

## 汉语句法分析的语料库方法研究..... 吴 升 (171)

|     |                       |       |
|-----|-----------------------|-------|
| 第1章 | 前言                    | (172) |
| 第2章 | 对若干基于语料库的句法分析工作的评述    | (173) |
| 2.1 | UCREL 基于成分似然语法的分析系统   | (174) |
| 2.2 | BSO 基于依存语法的句法分析器      | (175) |
| 2.3 | 宾夕法尼亚大学的广义互信息方法       | (176) |
| 第3章 | 汉语依存语法及其依存关系的定义       | (177) |
| 3.1 | 关于汉语依存语法模型中的几个重要概念的说明 | (178) |
| 3.2 | 汉语依存语法模型及句法分析的形式化描述   | (184) |
| 第4章 | 知识获取和句法分析器的建造         | (187) |
| 4.1 | 人工标注的形式及知识获取的过程       | (187) |
| 4.2 | 系统的组成及简介              | (188) |
| 4.3 | 句法分析器的主要处理流程及其策略      | (190) |
| 4.4 | 两个实例分析                | (190) |
| 4.5 | 关于实现中若干细节的说明          | (192) |
| 4.6 | 关于开放性语料处理的讨论及途径       | (194) |
| 第5章 | 总结与展望                 | (196) |
| 附录1 | 部分分析器的分析结果一览表         | (196) |
| 附录2 | 依存关系及代码一览表            | (198) |
| 附录3 | 部分义类类型及代码             | (200) |
| 附录4 | 词类类型及代码一览表            | (201) |
|     | 参考文献                  | (202) |

## 基于合一的汉语句法分析系统的研究与实现..... 栾 浩 (203)

|     |                   |       |
|-----|-------------------|-------|
|     | 前言                | (204) |
| 第1章 | 句法分析概论            | (204) |
| 1.1 | 句法分析技术的发展历史       | (204) |
| 1.2 | 句法分析技术的新发展        | (206) |
| 第2章 | UBCP 的基础语法理论和分析方法 | (206) |
| 2.1 | 基于合一的语法理论         | (206) |

|       |                          |       |
|-------|--------------------------|-------|
| 2.2   | Tomita 算法 .....          | (210) |
| 第 3 章 | UBCP 中存在的歧义问题及解决策略 ..... | (213) |
| 3.1   | 汉语的语法特征 .....            | (213) |
| 3.2   | 汉语分析中普遍存在着歧义问题 .....     | (213) |
| 3.3   | 相应的解决策略 .....            | (214) |
| 第 4 章 | UBCP 系统的基本设计思想 .....     | (218) |
| 4.1   | 句法分析的功能 .....            | (218) |
| 4.2   | 句法分析系统设计的一般原则 .....      | (219) |
| 4.3   | 分析器的基本设计思想 .....         | (220) |
| 第 5 章 | UBCP 的知识库 .....          | (222) |
| 5.1   | 电子词典 .....               | (222) |
| 5.2   | 语法规则库 .....              | (226) |
| 5.3   | 电子词典的控制机制 .....          | (229) |
| 5.4   | 规则库的控制机制 .....           | (230) |
| 第 6 章 | UBCP 系统的实现 .....         | (233) |
| 6.1   | 系统总体框图 .....             | (233) |
| 6.2   | 规则库处理子系统 .....           | (234) |
| 6.3   | 词典处理子系统 .....            | (234) |
| 6.4   | 用户接口子系统 .....            | (236) |
| 6.5   | 核心分析算法 .....             | (237) |
| 6.6   | 实例分析 .....               | (240) |
| 6.7   | 合一算法的讨论 .....            | (241) |
| 第 7 章 | UBCP 系统的实验结果和评价 .....    | (244) |
| 7.1   | 实验环境和实验结果 .....          | (244) |
| 7.2   | UBCP 系统评价 .....          | (244) |
| 结束语   | .....                    | (245) |
| 附录 1  | 本系统所用词语词类代码表 .....       | (246) |
| 附录 2  | 规则描写语言 DRDL 的书写规范 .....  | (246) |
| 附录 3  | 部分实验结果 .....             | (250) |
| 参考文献  | .....                    | (258) |

## 基于合一算法的汉语生成研究..... 李 东 (260)

|       |                    |       |
|-------|--------------------|-------|
| 第 1 章 | 绪论 .....           | (261) |
| 1.1   | 语言生成及其与理解的差别 ..... | (261) |
| 1.2   | 单句生成概述 .....       | (263) |
| 1.3   | 话语生成概述 .....       | (264) |
| 1.4   | 话语生成中的单句生成 .....   | (268) |
| 1.5   | 本文任务 .....         | (268) |

|      |                           |       |
|------|---------------------------|-------|
| 第2章  | 汉语句子意义的一种语用语义描述·····      | (269) |
| 2.1  | 一种句子意义的分析模式·····          | (269) |
| 2.2  | 句子意义的一种语用语义描述·····        | (271) |
| 2.3  | 小结·····                   | (278) |
| 第3章  | 汉语常用句式及其生成规律·····         | (278) |
| 3.1  | 汉语生成的特点·····              | (278) |
| 3.2  | 语义链·····                  | (279) |
| 3.3  | 小结·····                   | (285) |
| 第4章  | 生成系统设计·····               | (285) |
| 4.1  | 复杂特征集的引入·····             | (285) |
| 4.2  | 规则宜采用陈述性描述·····           | (287) |
| 4.3  | 基于合一的语法理论·····            | (287) |
| 4.4  | FUG 语法用于汉语生成·····         | (290) |
| 第5章  | 汉语句子生成实验系统的实现·····        | (295) |
| 5.1  | 数据结构·····                 | (296) |
| 5.2  | 系统中定义的几个框架处理函数·····       | (300) |
| 5.3  | 生成算法·····                 | (301) |
| 5.4  | 对生成中几个问题的探讨·····          | (302) |
| 5.5  | 结论·····                   | (307) |
| 第6章  | 话语生成实验系统的设计与实现·····       | (307) |
| 6.1  | 系统的总体设计·····              | (308) |
| 6.2  | 知识库信息的存储和管理·····          | (309) |
| 6.3  | 话语规划·····                 | (310) |
| 6.4  | 几个主要模块的程序实现·····          | (316) |
| 第7章  | 结束语·····                  | (316) |
| 附录1  | Fillmore 对英语在不同时期的格表····· | (318) |
| 附录2  | 汉语的格体系·····               | (318) |
| 附录3  | 汉语动词的态分类和态标记·····         | (322) |
| 附录4  | 名词语义体系(救火机器人话语生成系统)·····  | (324) |
| 附录5  | DGSS 的运行结果·····           | (324) |
| 参考文献 | ·····                     | (324) |

## 第2篇 汉字识别

|                 |                       |       |
|-----------------|-----------------------|-------|
| 印刷体汉字识别的研究····· | 朱夏宁                   | (328) |
| 第1章             | 引言·····               | (330) |
| 1.1             | 印刷体汉字识别研究工作的发展概况····· | (330) |

|     |                 |       |
|-----|-----------------|-------|
| 1.2 | 印刷体汉字识别的困难和识别方法 | (331) |
| 1.3 | 印刷体汉字识别的候选集确定   | (331) |
| 1.4 | 印刷体汉字识别的字识别     | (333) |
| 1.5 | 本章小结            | (339) |
| 第2章 | 印刷体汉字的特征选择      | (335) |
| 2.1 | 引言              | (335) |
| 2.2 | 印刷体汉字的特点        | (335) |
| 2.3 | 边框特征(部首)        | (335) |
| 2.4 | 局部特征(字根)        | (337) |
| 2.5 | 笔划端点和结点         | (337) |
| 2.6 | 本章小结            | (338) |
| 第3章 | 印刷体汉字结构特征的提取    | (338) |
| 3.1 | 笔划及其端点的提取       | (338) |
| 3.2 | 印刷体汉字的字根提取      | (342) |
| 3.3 | 印刷体汉字的部首提取      | (342) |
| 3.4 | 本章小结            | (346) |
| 第4章 | 印刷体汉字的结构统体特性及分析 | (346) |
| 4.1 | 印刷体汉字周边特征的分析    | (347) |
| 4.2 | 印刷体汉字游程统计特性及分析  | (349) |
| 4.3 | 印刷体汉字笔划特性及分析    | (350) |
| 4.4 | 印刷体汉字的字根统计特性及分析 | (352) |
| 4.5 | 印刷体汉字部首分布的统计及分析 | (353) |
| 4.6 | 本章小结            | (357) |
| 第5章 | 实验性印刷体汉字识别系统    | (357) |
| 5.1 | 系统设计的一些考虑       | (357) |
| 5.2 | 系统的组成与工作过程      | (358) |
| 5.3 | 实验结果            | (360) |
|     | 结束语             | (362) |
|     | 参考文献            | (363) |

## 人工神经网络印刷汉字识别系统的研究..... 徐 宁 (365)

|     |                 |       |
|-----|-----------------|-------|
| 第1章 | 绪论              | (367) |
| 第2章 | 汉字识别的嵌套分层结构     | (368) |
| 2.1 | 印刷体汉字的特点        | (368) |
| 2.2 | 汉字识别的嵌套分层结构     | (369) |
| 2.3 | 本章小结            | (371) |
| 第3章 | 基于人工神经网络的汉字特征提取 | (372) |
| 第4章 | 汉字识别自组织聚类人工神经网络 | (376) |

|                               |                      |           |
|-------------------------------|----------------------|-----------|
| 4.1                           | 引言                   | (376)     |
| 4.2                           | 汉字识别自组织聚类人工神经网络      | (377)     |
| 4.3                           | 实验结果与分析              | (381)     |
| 4.4                           | 本章小结                 | (386)     |
| 第5章                           | 汉字细分类BP网络            | (386)     |
| 5.1                           | 汉字细分类BP网络结构          | (386)     |
| 5.2                           | BP算法的改进              | (386)     |
| 5.3                           | 细分类网络训练样本的选择         | (387)     |
| 5.4                           | 本章小结                 | (389)     |
| 第6章                           | 基于协作模块神经网络的汉字识别方法    | (389)     |
| 6.1                           | 问题的提出                | (389)     |
| 6.2                           | 协作模块神经网络             | (390)     |
| 6.3                           | 结论与实验结果              | (393)     |
| 第7章                           | 神经网络多体印刷汉字识别实验系统的实现  | (393)     |
| 7.1                           | 系统的组成和建立             | (394)     |
| 7.2                           | 实验结果                 | (396)     |
|                               | 参考文献                 | (397)     |
| <b>联机手写汉字识别和印刷体汉字识别的研究与实践</b> |                      | 杨德顺 (401) |
| 第1章                           | 联机手写汉字识别             | (402)     |
| 1.1                           | 联机手写汉字识别方法概述         | (402)     |
| 1.2                           | 手写汉字模式的形式化描述—汉字表达式   | (404)     |
| 1.3                           | 笔划有序列生成              | (407)     |
| 1.4                           | 汉字表达式的机器学习           | (410)     |
| 1.5                           | 联机手写汉字识别系统的实现        | (414)     |
| 第2章                           | 印刷体汉字识别              | (418)     |
| 2.1                           | 汉字特征点的定义及性质          | (418)     |
| 2.2                           | 汉字特征点的自动抽取           | (419)     |
| 2.3                           | 基于特征点的印刷体汉字识别模拟系统    | (423)     |
| 附录1                           | 局部结构分析法抽取笔划特征点的步骤示例  | (429)     |
| 附录2                           | 关键背景点抽取的步骤示例         | (434)     |
| 附录3                           | 局部结构分析法抽取笔划特征点的稳定性示例 | (437)     |
| 附录4                           | 局部结构分析法抽取笔划特征点程序运行实例 | (438)     |
|                               | 参考文献                 | (442)     |
| <b>手写印刷体汉字识别的研究</b>           |                      | 吴智彪 (444) |
| 第1章                           | 手写印刷体汉字识别概述          | (445)     |
| 1.1                           | 手写印刷体汉字识别的发展         | (445)     |

|     |                        |       |
|-----|------------------------|-------|
| 1.2 | 识别手写印刷体的结构相似法          | (447) |
| 第2章 | 粗分类特征的选择和提取            | (448) |
| 2.1 | 粗分类特征的选择               | (48)  |
| 2.2 | 汉字图象的多角形近似             | (450) |
| 2.3 | 实现投影的扫描线算法             | (451) |
| 2.4 | 四边框形状特征码的获取            | (451) |
| 2.5 | 本章小结                   | (452) |
| 第3章 | 粗分类的实现                 | (453) |
| 3.1 | 特征库的结构                 | (453) |
| 3.2 | 搜索算法                   | (454) |
| 3.3 | 粗分类结果的获取               | (454) |
| 3.4 | 本章小结                   | (455) |
| 第4章 | 细分特征抽取——笔划段快速合并法       | (456) |
| 4.1 | 细化算法及多角形近似抽取笔划法的不足     | (456) |
| 4.2 | 近似多角形的特点               | (456) |
| 4.3 | 合并算法                   | (457) |
| 4.4 | 本章小结                   | (458) |
| 第5章 | 一种新的图匹配法——相关属性关系图启发式匹配 | (458) |
| 5.1 | 常用关系结构图匹配方法的问题         | (459) |
| 5.2 | CARG 的数学描述及启发式匹配算法     | (460) |
| 5.3 | 手写印刷体汉字图匹配的实现          | (462) |
| 5.4 | 匹配实例                   | (465) |
| 5.5 | 本章小结                   | (467) |
|     | 参考文献                   | (467) |

## 手写汉字识别的研究 ..... 曾棋荣 (470)

|     |                 |       |
|-----|-----------------|-------|
| 第1章 | 引言              | (471) |
| 1.1 | 手写汉字识别的进展       | (471) |
| 1.2 | 手写汉字识别研究方向      | (473) |
| 1.3 | 本论文的主要工作        | (474) |
| 第2章 | 特征的选择和抽取        | (474) |
| 2.1 | 手写体汉字的特点        | (474) |
| 2.2 | 特征的选择与抽取        | (475) |
| 第3章 | 粗分类及机器学习        | (479) |
| 3.1 | 粗分类             | (479) |
| 3.2 | 快速机器学习          | (479) |
| 第4章 | 手写体汉字识别中的特征匹配判别 | (480) |
| 4.1 | 特征匹配判别的一般方法及问题  | (480) |