

焊接 CAPP 中数据挖掘技术的研究

摘 要

近年来,随着计算机应用技术的迅速发展及其在生产领域的普及,国内外开展了许多关于焊接专家系统,焊接数据库,焊接 CAPP 系统的研究工作。由于焊接工艺直接影响到产品的质量,对产品的安全起着决定性作用,而焊接工艺设计过程本身的复杂性和多因素性,人们需要引入数据库技术对焊接工艺文件和资料进行保存和管理。面对庞大的数据量,如何将生产数据进行整理、分析,找出对企业有价值的潜在工艺信息,一直是困扰企业的难题。人们希望能够对焊接过程中产生的数据进行科学分析,发现内在规律,得出科学的预测结果,以便更好地利用这些数据来指导生产活动。数据挖掘技术被认为是目前解决“数据爆炸”和“数据丰富,信息贫乏”的一种有效方法。鉴于以上原因,本课题针对具有代表性的汽车行业的车身焊接工艺设计活动为背景,应用数据挖掘技术,研究和开发出一种更为智能化、集成化、可重构的创成式焊接 CAPP 系统。

本文首先阐述了焊接 CAPP 系统的发展现状及研究意义,指出发展中的不足,以此为依据介绍了课题研究的目的,提出了基于数据挖掘技术的焊接 CAPP 设计思想。

文章简单介绍了数据挖掘技术的基本理论及常用方法,并从经典的关联规则算法和决策树算法角度进行详细介绍,针对课题研究内容提出关联规则算法的改进算法。

在建立数据挖掘模块原理的基础上,根据焊接工艺设计的原理和特点,结合企业的实际需求和积累多年的焊接工艺数据,提出了系统的设计思想,确定了系统总体框架及模块功能,重新制定了设计、开发方案,实现了焊接工艺设计和数据挖掘技术的结合。

最后,对焊接 CAPP 系统各个模块进行详细介绍,同时对系统开发的数据挖掘模块进行实例演示,验证系统的正确性和性能的实用性。

作者根据实际需要开发的基于数据挖掘技术的创成式焊接 CAPP 系统,有效地减轻了工艺人员的工作强度,缩短了工艺设计的周期,提高工艺设计的质量和效率,使焊接技术人员能有更多的时间、精力进行新技术、新工艺的研发工作,同时提高了焊接厂家的现代化管理水平和科技实力,增强了企业的竞争能力,在保持同行业领先的基础上赶超国际先进水平。系统关于数据挖掘模块的研究和开发,对于我国焊接行业实现计算机辅助工艺设计,具有较高的理论意义和实际应用价值。

关键词: 焊接; 工艺设计; CAPP; 数据挖掘; 关联规则

Data-mining Technology investigated in Welding CAPP System

Abstract

Along with the rapid development of computer application technology and its permeation in manufacturing domain in recent years, there have been a lot of institution work about welding expert system, welding database, welding computer aided process planning in and abroad of China. As welding technic has an influence to the quality of the products directly and it determinate the safe of products, and the process of welding technic design is complicated and it has too many diathesis itself, people should introduce database technology to save and manage welding technic documents and information. Affronting the large amount of data, how to tidy up, analyz these producing data, find out the potential information which is valuable to companies, are difficult problems which always puzzle companies. People want to direct production process from analyzing these data in welding production, finding ingenerate regulation and getting scientific prediction results. Data mining is considered to be an effectively method to solve “data boost” and “data abundance and information indigence” at the present time. Whereas all these reasons, this paper using data-mining technology investigate and develop a more intelligent, composite, reconstructed generated welding CAPP system based on welding technic design.

Firstly, the paper expatiates current develop status and investigative meaning, points out the lack of CAPP. Then it describes the purpose of research and lodges the designing thoughts of welding CAPP based on data-mining.

This paper introduces basic theories and common methods of data-mining briefly, then introduce classical related regulation algorithm and decision trees in details. Aiming at the content of investigation task, this paper also introduces the improved algorithm of related regulation.

Based on the elements of data-mining module, according to the theory and characteristic of welding design, combining the requirement of company in practice and welding technic data which have been accumulated these years, this article lodge the thought of design, confirm the frame and module function of the system collectivity, establish design and exploitation scheme again, realize the combination of welding technic design with data-mining.

At last, this article introduces all the modules of the CAPP system, meanwhile, it demonstrates all the developed modules, validate the correctness of system and

the practicability of capability.

The developed generated welding CAPP system using data-mining technology, which can reduce the intension of work, shorten cycle of technic design, enhance the quality and efficiency of designing process, make people have more time and energy to develop new technology and technic. It can also improve the modern management level and science strength, enhance the competition ability of company and overtake international advanced level. The institution and development of data-mining module in this system, has upper theory meaning and application value in practice to welding field using computer aided process planning in our country.

Key words: weld; technic design; CAPP; data-mining; related regulation

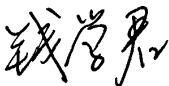
插图清单

| | |
|-------------------------------------|----|
| 图 1-1 CAPP 系统基本结构 | 2 |
| 图 3-1 创成式 CAPP 系统原理图 | 23 |
| 图 3-2 基于数据挖掘技术的创成式 CAPP 系统原理图 | 24 |
| 图 3-3 数据库设计的基本过程 | 28 |
| 图 3-4 观焊接工艺数据库 E-R 模型图 | 31 |
| 图 3-5 系统数据库结构图 | 33 |
| 图 3-6 数据库组件体系结构 | 35 |
| 图 4-1 系统软件开发流程 | 36 |
| 图 4-2 系统封面 | 37 |
| 图 4-3 登陆窗口 | 38 |
| 图 4-4 系统用户界面 | 39 |
| 图 4-5 母材化学成分查询结果界面 | 40 |
| 图 4-6 母材力学性能查询结果界面 | 41 |
| 图 4-7 母材组织性能查询结果界面 | 41 |
| 图 4-8 焊条力学性能查询 | 42 |
| 图 4-9 焊丝化学成分查询 | 43 |
| 图 4-10 保护气体特性用途查询 | 43 |
| 图 4-11 熔炼焊剂特性用途查询 | 44 |
| 图 4-12 烧结焊剂特性用途查询 | 44 |
| 图 4-13 坡口工艺查询 | 45 |
| 图 4-14 坡口工艺设计数据库维护 1 | 46 |
| 图 4-15 坡口图片预览 | 46 |
| 图 4-16 坡口工艺设计数据库维护 2 | 47 |
| 图 4-17 详细信息示意图 | 47 |
| 图 4-18 测试数据选择界面 | 48 |
| 图 4-19 结果集 R1 | 48 |
| 图 4-20 中间项集 M1 | 49 |
| 图 4-21 中间项集 MM1 | 49 |
| 图 4-22 数据源 S2 | 50 |
| 图 4-23 结果集 R2 | 50 |
| 图 4-24 详细信息示意图中间集 M2 | 51 |
| 图 4-25 数据源 S3 | 51 |
| 图 4-26 结果集 R3 | 52 |
| 图 4-27 支持度计算 | 52 |

| | |
|---------------------------|----|
| 图 4-28 置信度计算 | 53 |
| 图 4-29 焊接工艺参数的支持度计算 | 53 |
| 图 4-30 焊接工艺参数的置信度计算 | 54 |

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 合肥工业大学 或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

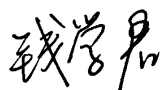
学位论文作者签名：

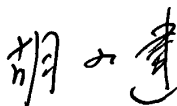
签字日期：2009年4月19日

学位论文版权使用授权书

本学位论文作者完全了解 合肥工业大学 有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权 合肥工业大学 可以将学位论文的全部或部分内 容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密的学位论文在解密后适用本授权书)

学位论文作者签名：

导师签名：

签字日期：2009年4月19日

签字日期：2009年4月19日

学位论文作者毕业后去向：

工作单位：

电话：

通讯地址：

邮编：

致 谢

本文是在导师胡小建教授的悉心指导与关怀下完成的。导师严谨的治学态度、渊博的学识、敏锐的科学思维、大胆创新的科学精神以及无私奉献的高尚情操、乐观大度的生活态度令学生终生难忘。三年来，胡老师不仅在学业上给予我极大的指导和教诲，而且在生活上给我无微不至的关怀和帮助；不仅传授科研治学的的思路和方法，更注重教导为人处世的道德和原则。值此论文完成之际，谨向恩师致以最崇高的敬意和最诚挚的感谢。

将近三年的硕士学习，让我不但在专业理论知识上有所收获，更在人际关系、为人处世上学到了很多。这些收获来自于关心指导我的老师，来自于一直关爱支持我学习的家人和亲友，来自于实验室的各位研究生同学，来自于所有一同努力一同成长的朋友。

论文工作期间，我还要特别感谢焊接教研室的徐道荣老师、李萌盛老师，感谢他们在研究工作中的热情指导和提出的宝贵建议。同时对焊接教研室的其他老师一并表示真诚的感谢，此外，还要感谢材料学院焊接实验室的老师在试验过程中给予的大力的支持，正是有了他们的指导和帮助，我的试验才能顺利完成。

最后感谢在研究生学习期间给予支持和帮助的朋友，他们是李旭光、王丽婷、仝维、卜师瑞、吴志国、邵光辉、吴丽娜.....。

在毕业之际，对所有关心和陪伴我的老师与朋友的祝福是无法用简简单单的几句话能表达的，此时无声胜有声，祝愿大家一生平安幸福。

作者：钱学君

2009年4月

第一章 绪论

制造是人类文明的基石，是人类得以生存与发展最基本、最重要的手段之一。信息技术的出现与发展，为发展中国家快速实现工业化带来了新的机遇。实现工业生产和管理的计算机化，成为提高工业生产率及工艺设计效率，降低产品制造成本和生产周期，提高企业市场竞争力的关键因素之一^[1]。

在现代化工业生产中，产品更新换代的速度越来越快，传统的生产方式已经不能适应现代工业化生产的发展需要。近年来，计算机技术及信息产业的迅猛发展，利用计算机来辅助生产成为必然，产品开发与投产也在不同程度上实现了计算机化，相应产生了计算机辅助设计（CAD）、计算机辅助工艺设计（CAPP）和计算机辅助制造（CAM）。其中，计算机辅助工艺设计（Computer Aided Processing Planning, CAPP）不仅是联接 CAD 和 CAM 的纽带，也是先进制造技术的核心技术之一，是实现 CIMS、并行工程（Concurrent Engineering, CE）乃至敏捷制造（Agile Manufacturing, AM）的重要基础性技术^[2]。

但是，同 CAD、CAM 等计算机辅助技术相比，CAPP 在应用方面仍然是薄弱环节。一方面，工艺设计问题的复杂性和多因素性决定了 CAPP 系统开发面临着工艺知识获取难，企业工艺信息模型与资源状况迥异的现实；另一方面，受 CAPP 发展的大背景和指导思想影响。因此，尽管国内外 CAPP 的深入研究已经持续进行了 30 多年，各种新概念、新方法不断涌现，CAPP 的发展缺乏坚实的实践基础是公认的事实。其中一个重要问题是传统意义上 CAPP 仅完成工艺设计，而在实际生产中，工艺设计只是工艺信息生成和处理全过程中的一部分工作，并没有涉及大量烦琐的工艺信息管理工作，开发出的系统离实用化的目标相去甚远。工艺设计的局部高效率并没有带来工艺信息管理整体工作的高效率，因而使 CAPP 系统的功能和效率大打折扣。近年来，在先进制造技术发展的带动下，对 CAPP 的理解已经完全突破了其字面含义，CAPP 系统正在向设计与管理一体化的制造工艺信息系统的方向发展^[3]。以下从 CAPP 的基本原理、工作原理，以及国内外现状、发展趋势等方面进行介绍。

1.1. CAPP 技术概述

1.1.1 CAPP 系统原理

计算机辅助工艺设计（CAPP-Computer Aided Process Planning）是适应多变的 market 环境和需求多样化发展起来的一门机械制造应用系统。它源于成组技术（GT）在工艺设计中的应用，是一种集管理、文字编辑、智能化生成为一体的工艺软件系统^[4]。在传统的 CAPP 研究中，依据工艺决策方式，将 CAPP 系统划分为两大类：修订式（variant, 亦称派生式）CAPP 系统和创成式（generative, 亦称生成式）CAPP 系统。根据技术发展及实际开发需要，也有兼容上述两种方法的混合式系统，以及影响更大的应用人工智能（AI）及专家系统（ES）技

术的 CAPP 专家系统。

1.1.2 CAPP 系统的基本结构

CAPP 系统的基本结构如下图 1-1 所示，包括信息输入、输出、数据库、工艺决策等^[5]。工艺决策以规则为基准，进行必要的比较、计算和决策，生成工艺规程。CAPP 系统的组成与其开发环境、产品对象及规模大小等有关。

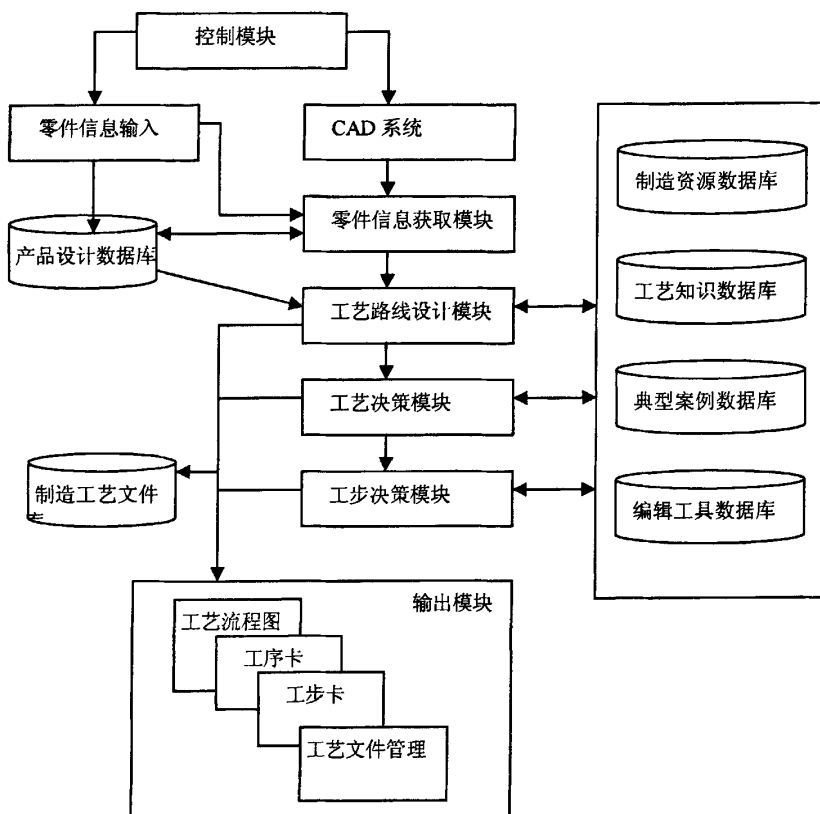


图 1-1 CAPP 系统基本结构

控制模块：协调各模块的运行，实现人机之间的信息交流，控制产品设计信息获取方式。

零件信息获取模块：用于产品设计信息输入。

工艺过程设计模块：进行加工工艺流程的决策，生成工艺卡。

工序决策模块：选定加工设备、定位安装方式、加工要求，生成工序卡。

工步决策模块：选择刀具轨迹、加工参数，确定加工质量要求，生成工步卡及提供形成 NC 指令所需的刀位文件。

输出模块：输出工艺流程卡、工序和工步卡，工序图等各类文档。

产品设计数据库：存放有 CAD 系统完成的产品设计信息。

制造资源数据库：存放企业或车间的加工设备、工装工具等制造资源的相

关信息。

工艺知识数据库：用于存放产品制造工艺规则、工艺标准、工艺数据手册、工艺信息处理的相关算法和工具等。

典型案例库：存放各零件族典型的零件的工艺流程图、工序卡、工步卡、加工参数等数据，供系统参考使用。

编辑工具库：存放工艺流程图、工序卡、工步卡等系统输入输出模板、手工查询工具和系统操作工具集等。

制造工艺数据库：存放由 CAPP 系统生成的产品制造工艺信息，供输出工艺文件、数控加工编程和生产管理与运行控制系统使用。

1.1.3 CAPP 系统的基本类型

CAPP 系统就其工作原理可以分为以下几种。

(1) 检索式 CAPP 系统

在检索式 CAPP 系统中，设计好的零件标准工艺被编号，并存储在计算机中；在制定零件的生产工艺时，可根据输入的零件信息进行搜索，查找合适的标准工艺。可见，检索式 CAPP 系统简单实用，但是由于标准工艺为数有限，大量的零件不能被覆盖，所以应用范围有限。

(2) 派生式 CAPP 系统

派生式 CAPP 系统，根据成组技术（Group Technology）原理将零件划分到不同的零件组，按零件组编制出标准工艺，并将其存储到计算机中；在为新零件设计工艺时，输入该零件的成组技术代码，检索到相应零件组的工艺，而后根据该零件的特点，由计算机自动进行工艺参数的修改，从而获得所需的工艺。由派生法产生的工艺往往需要在成组技术和数据库技术的基础上，利用零件的相似性，通过对零件族主样件的典型工艺进行修改，从而获得所需的工艺。

(3) 创成式 CAPP 系统

早期的创成式 CAPP 系统将决策逻辑植入程序代码中，依靠决策树和决策表进行设计。这种创成方法难以随零件类型和生产环境的变化自动生成合理的工艺规程。当前，创成式 CAPP 系统已发展成智能化 CAPP 系统，也称为 CAPP 专家系统，它以推理+知识为特征，将工艺实践经验以生产式规则或其他知识表达形式纳入知识库中，成为系统推理机的依据。系统推理机和知识库实现了分离，便于系统开发人员和工艺人员的合作。工艺过程设计的主要问题不是数值计算，而是对工艺信息和工艺知识的处理，这正是专家系统的特长所在。创成式 CAPP 系统具有较高的柔性，适应范围较广，而且便于与 CAD 和 CAM 集成^[6]。

1.2 CAPP 系统的发展历程

工艺过程设计是企业各部门信息汇集的重要环节，生产管理和设计部门依靠 CAPP 系统不仅可以对产品设计、工艺设计、产品制造、产品测试等各个阶段提供全面支持，缩短工艺设计周期，保证工艺设计质量，而且可以提高产品工艺的继承性，最大限度地利用现有资源，提高产品的市场竞争力。

作为工业生产中最重要材料成型方法之一，焊接的应用遍及了航天、造船、化工、电力、建筑、汽车、微电子等领域，其质量和可靠性直接关系到最终产品的性能与安全，其成本也在较大程度上影响到产品的最终成本。焊接工艺的优劣直接影响产品的质量、制造成本、生产效率等，同时焊接零件的产品系列及型号非常多，工艺又相当复杂，依靠焊接工艺人员手工处理非常麻烦。因此，将计算机辅助工艺设计应用于焊接工艺设计中，可以帮助工艺人员选定适用的焊接工艺规程，进行焊接工艺的规范设计，节能降耗，是提高企业经济效益和市场竞争力的重要保证。在生产过程中，CAPP 的应用可以将工艺设计人员从大量繁重的、重复性的手工劳动中解放出来，将主要精力转向新产品的开发、工艺装备的改进及新工艺的研究等创造性的工作，有助于推动企业开展工艺设计标准化和最优化工作。因此，CAPP 受到焊接领域的重视，以 CAPP 代替传统方式进行工艺设计和管理得到了大的发展。

1.2.1 国外 CAPP 系统的发展过程

60 年代末，人们就开始了计算机辅助工艺规划（CAPP）的研究与开发，第一个 CAPP 系统是挪威 1969 年推出的 AUTOPROS（自动工艺设计）系统，1973 年正式将 AUTOPROS 系统商品化。但在 CAPP 发展史上具有里程碑意义的是设在美国的国际性组织 CAM-I 于 1976 年开发的 CAPP（CAM-I's Automated Process Planning）系统。70 年代以来，CAPP 系统在世界各国均有所开发，如日本日立制作所的 HIMTCS 工艺设计、美国的 MIPLAN 系统、国际机械制造研究会（CIRP）的 CAPP 系统等。随着计算机及其相关技术的发展，各种 CAPP 系统在发展中不断改进提高和互相渗透，从 20 世纪 80 年代开始，各国将研究重点放在将人工智能（AI）、专家系统、人工神经网络、模糊推理、基于实例的推理等技术在 CAPP 系统的应用中，以及对混合式 CAPP 系统的开发^[7]。1985 年和 1986 年，美国机械工程师协会（ASME）连续两年召开 CAPP 的学术研讨会，随后的几年中，国际生产工程研究会（CIRP）连续两次召开 CAPP 专题讨论会。目前，国外各类生产制造自动化学术会议、计算机集成制造技术学术会议等都设有 CAD、CAPP、CAM 或人工智能技术在 CAPP 中的应用专题。

1.2.2 国内 CAPP 系统在焊接领域的应用概况

在国外，CAPP 的发展已有 30 多年的历史，国内也进行了近 20 年的研究与实践。焊接工作者早在 20 世纪 80 年代后期就在焊接工艺设计与管理中引进了 CAPP 技术。1990 年前后，可以认为是国内焊接领域 CAPP 发展的起步时期，从 1988 年开始，太原重机厂在长城 GW 系统机上自行开发了人机交互式计算机辅助焊接工艺规程设计软件系统 TZ-WCAPP，该系统以成组技术为基础，以专家系统思想为指导，采用检索法与创成法相结合，可以认为是一种综合式 CAPP 系统，该系统还应用集成概念，实现了 CAD、WCAPP、MRP 的一体化。1994 年左右，为数众多的企业开始研究焊接工艺 CAPP 系统，清华大学与大连起重机器厂合作开发了 CSCAMP 系统；第一重型机械集团公司开发了基于网络的 WTPMIS 系统；哈尔滨工业大学与哈尔滨锅炉厂合作开发了 PQRDBMS 系统。这一时期的系统一般使用 BASIC 或 VB、FOXPRO 开发，运行在 Windows 上。少数系统具备一定的工艺设计自动化功能，绝大多数系统主要进行焊接工艺的管理，但是提供的功能比以前的系统要强很多，界面也更加友好。近年来，已有的 CAPP 系统得到进一步发展，在 AutoCAD 等自主图形开发平台上进行二次开发的 CAPP 系统得到了普及应用，适用性大大加强，有的系统还实现了网络化^[8-9]。

1.3 焊接领域新一代焊接 CAPP 系统

1.3.1 焊接 CAPP 系统的确定

所谓焊接 CAPP 系统，是将 CAPP 系统与企业生产应用的焊接工艺有机结合，使得焊接工艺人员从大量重复单一的技术准备工作中解放出来，进而从事更具有创造价值的工艺研究；同时可以将过去成熟可靠的工艺文件放入数据库中，便于技术人员选用；另外也有利于保证工艺文件的正确性，统一性的要求。

1.3.2 焊接 CAPP 系统构成

焊接 CAPP 系统由 6 个模块组成^[10]：

- (1) 焊接信息输入模块：该模块主要是用来实现焊接信息的输入；
- (2) 工艺过程生成模块：该模块用来完成工艺决策问题，包括焊接方法的选择，焊接顺序确定等；
- (3) 工艺规程生成模块：该模块主要是生成焊接材料、焊接规范等参数；
- (4) 工艺卡片生成模块：主要完成工艺卡片的生成；
- (5) 工艺文件编辑与管理模块：对已生成的工艺文件进行编辑和管理，以求生成用户所需要的标准准确的工艺文件；
- (6) 工艺文件输出打印模块：用来完成有关资料的汇总及各类工艺文件的输出。

1.3.3 焊接 CAPP 系统的特点及实施效果

(1) 人机界面灵活，可以在多个工作窗口内自由切换，任意选择横向或竖向编辑方式；

(2) 采用关系型数据库，将各种格式的工艺卡保存在数据库中，设计人员可以方便地进行添加、删除、检索、浏览、打印工艺文件等操作；

(3) 在焊接工艺人员设计工艺参数时，为用户提供参考数据和提示信息，这些数据和信息来源于国标、企标或经验，用户在设计过程中可以修改这些数据；

(4) 实用性强，大量的数据编辑采用自动查表方式选择输入，能自动生成相关的焊接工艺文件报表，并提供多种灵活的数据输出方式，如 EXCEL 表格文件，纯文本文件，SQL 数据库文件等；

(5) 自动汇总零件清单，典型工艺路线、工艺装备清单、工艺装备一览表等文件，汇总焊接材料消耗定额，产生标准件的材料消耗定额率；

(6) 焊接工艺人员通过焊接 CAPP 系统与设计部门进行动态信息交流，推进 CAPP 系统工艺标准化、规范化的工作，确保工艺设计的准确性。

(7) 基于局域网，具有完善的权限管理功能。

通过 CAPP 系统的应用和实施，焊接工艺人员可以根据及时提交的工艺数据帮助相关部门准确地制定决策，提高工作效率，促进了工艺的标准化建设工作，提高工艺设计水平和质量；同时积累大量的工艺数据，将企业的信息化水平推上一个新的高度^[9]。

1.4 焊接 CAPP 系统的发展趋势

随着国家对制造业信息化政策的制订和落实以及制造技术的发展，对 CAPP 在焊接中的应用提出了更高的要求。一个实用可行的焊接 CAPP 系统，必须能够优化工艺设计，提高标准化水平，指导生产过程与生产管理，因此，无论在技术上还是在系统上都需要对 CAPP 的发展进行深入的研究，焊接 CAPP 的发展趋势主要包括以下几点：

1.4.1 人工智能技术的应用

焊接 CAPP 系统不会停留在以解决事务性、管理性工作为主的阶段，除了作为工艺设计辅助工具，还有将工艺专家的经验 and 知识积累起来加以充分利用的任务。人工智能技术在系统中通过符号处理和推理机制实现逻辑推理能力，提高系统的适应性和通用性。在原有 CAPP 的开发模式、体系结构框架内，结合现代计算机、信息等相关技术的进展，采用新的决策算法、发展新的功能，并已在并行、分布、平台、面向对象等方面进行着有意义的尝试。

1.4.2 面向产品的 CAPP 工具和应用平台

各企业的工艺环境、管理模式千差万别，为了能使焊接 CAPP 系统在企业中能更好地应用推广，提高通用性，就需要 CAPP 系统提供更好的开发模式。应用面广、适应性强的平台型焊接 CAPP 系统，已经成为开发和应用的趋势。将 CAPP 系统的功能分解成一个个相对独立的工具，工艺人员可以根据企业具体情况输入数据和知识，形成面向特定制造和管理环境的焊接 CAPP 系统，并且可以进行二次开发。

1.4.3 研究支持全面集成的工艺规划方法

随着企业信息化建设的不断深入，CAD、CAM、CAE、MIS、PDM、ERP 等各种单元系统在企业均得到很好的应用，但由于单元系统之间的信息得不到充分利用，很多信息需要重新输入，面向单个单元系统的 CAPP 已远远不能满足企业的整体发展需要了，所以近年来面向产品的集成化 CAPP 系统研究显得尤为重要。在并行工程思想的指导下实现 CAD、CAM、CAPP 等系统的全面集成，进一步发挥 CAPP 在整个生产活动中的信息中枢和功能调节作用。计算机集成制造是现代化制造业的发展趋势，作为集成系统中的一个单元技术，CAPP 系统集成化也是必然的发展方向^[11]。

1.4.4 兼顾专业化和通用化的发展方向

通用化是目前焊接 CAPP 研究的一个重要方向，通用化使得软件品质得以保证，有助于加速 CAPP 系统及其相关技术的标准化，并且对 CAPP 系统在企业中的应用推广具有重要意义。在专业化方面，针对焊接工艺条件复杂多变，工艺内容与生产实际环境联系非常紧密的特殊性，积极摸索焊接加工工艺的规律，与企业特定应用环境牢牢结合在一起，形成专业化的工艺推理和决策模型，才有可能在当前的人工智能技术条件下，为工艺人员提供较实用的智能化工艺决策方案^[12]。

1.4.5 CAPP 系统的柔性和集成性并重

CAPP 系统的柔性是为了满足企业快速变化的需求，一方面市场的快速变化要求企业迅速做出反应，利用包括 CAPP 系统在内的 CIMS 更改产品的设计或者开发新产品，另一方面企业内部规章制度办公环境的变动要求 CAPP 系统与之相关的内容必须快速更新。CAPP 系统的集成性是 CAPP 有效性的前提，CAPP 与其它系统的集成是通过数据交互实现的，因此保证 CAPP 的集成性就必须满足特定的数据格式要求。这就要对与 CAPP 进行集成的系统数据格式进行分析，并分别提供不同的接口。通过在同一数据源上提供对应不同系统（CAD/CAM）的多种接口，同时保证接口的可扩充性，就可以同时满足集成

性和柔性的要求。当然，这会给软件的开发带来一定的难度。

1.4.6 适应先进生产模式的焊接 CAPP 系统向网络化方向发展

网络化是系统集成应用的必然要求，CAPP 对内实现各种角色、工种的并行工艺设计，对外与 CAD 实现双向数据交换，与 CAQ、CAM、PDM 等的集成应用都需要网络技术支持才能实现企业级乃至更大范围的信息化。在全球经济的趋势下，要组织好生产加工，就必须借助于网络技术，构建面向动态联盟企业的 CAPP 系统。CAPP 系统结构是企业内 CAPP 系统结构的扩展，也是动态联盟企业的 CAPP 系统的需求。企业的动态联盟性以及网络系统的安全性等问题都是基于网络的 CAPP 系统所必须考虑的，目前各方面的技术发展和成熟为基于网络的 CAPP 系统开发奠定了很好的基础^[13-14]。

1.5 现有焊接 CAPP 系统存在的不足

在过去的十多年中，国内焊接领域对 CAPP 系统的研究取得了很大的成绩，许多企业实现了应用计算机编制和管理焊接工艺，在提高产品质量和经济效益的同时，也向 CIMS 迈出了第一步，但还存在以下不足^[15-18]：

1.5.1 CAPP 研究与开发目标的片面性

长期以来，焊接 CAPP 系统一直是开发代替工艺人员的自动化系统，过分强调了工艺决策的自动化，在此目标指导下开发出的自动化 CAPP 系统，虽融入了专家的知识 and 经验，但在运行时通常需要用户按规定描述方法交互输入零件信息和加工环境配置信息，然后由系统进行自动决策，这就导致了系统的开发周期长，费用高，难度大，使用人员很难参与，适用性差；工艺人员难以掌握系统的使用方法，而且在使用中仍需交互输入大量信息，麻烦而又易出错；工艺设计与管理不能集成，不仅工作效率低，而且很难保证工艺信息和文件的准确性、一致性。

1.5.2 通用化程度低，应用范围过于狭窄

由于焊接 CAPP 系统涉及问题的复杂性和对企业具体环境的依赖性，每一种系统只适用于少数几种零件，对焊接 CAPP 系统的研究与开发常常针对特定制造对象、特定制造环境甚至特定设备进行，系统功能和应用范围有限，缺乏适应生产环境变化的灵活性和适应性，从而造成开发、实施具体应用系统困难，难以满足先进制造系统对焊接 CAPP 系统的使用和发展需求。

1.5.3 对 CAPP 的系统化、实用化、工程化研究重视不够

尽管各种新概念、新方法和新技术在焊接 CAPP 系统中不断获得应用，但

主要是针对各种决策功能，人们始终缺乏对 CAPP 的系统化、实用化和工程化的研究，缺乏各种技术之间及技术管理与人之间的集成，使新技术、新概念和新方法的采用对 CAPP 系统的发展所起作用甚小，也使得焊接 CAPP 系统的整体发展缺乏坚实的实践基础和动力。

1.5.4 注重表现形式，对互联网及大量的离线工艺数据重视不够

21 世纪是知识经济占主要地位的世纪。针对互联网及离线的焊接数据库数据量较大，而在实际生产过程中，部分焊接 CAPP 系统着力解决工艺设计本身的问题，却忽略了工艺信息的管理。多数焊接 CAPP 系统只能完成单纯的工艺文件管理，工艺设计过程的随意性比较大，很难及时掌握项目组成员的任务完成情况，使得工艺信息得不到充分的利用，信息化建设受到限制，对此应该引起人们的深思。

1.6 课题研究内容及意义

近年来，随着计算机技术的发展及其在各领域应用的普及，国内外也开展了许多焊接专家系统，焊接数据库，焊接工艺 CAPP 等研究工作。CAPP 系统在企业中已经获得了广泛的应用。在所开发的 CAPP 系统中，有些系统已得到了实际应用，但从应用角度来看，CAPP 系统必须能快速地、柔性地响应产品和需求的变化，缩短工艺设计周期，提高开发效率，并实现与其它制造环节的集成。目前的焊接 CAPP 系统还不能满足上述要求。

面对庞大的数据量，如何将生产数据进行整理、分析，找出对企业有价值的信息，却一直是困扰企业的难题。尽管在国内外的焊接 CAPP 研究中，各种新概念、新方法不断涌现，但由于涉及面太广，具体的针对性不强，大多数还是存在着实用功能或实用性不强等缺点。人们希望能够对焊接过程中产生的数据进行科学分析，发现深层次的、手工无法发现的内在规律，得出比较科学的预测结果，以便更好地利用这些数据来指导生产活动。

数据挖掘技术被认为是目前解决“数据爆炸”和“数据丰富，信息贫乏”的一种有效方法。挖掘数据库中蕴涵的知识是目前 CAPP 研究的热点，也是 CAPP 系统进一步深化应用的必然趋势。在国外，数据挖掘技术已经广泛地应用于金融业、零售业、远程通讯业、政府管理、制造业等信息化程度较高的行业，而国内则主要集中在算法学习以及有关数据挖掘理论方面的研究，将数据挖掘技术应用于焊接领域实际生产过程的还为数不多。

数据挖掘技术的研究将会极大地推动知识创新，提出将数据挖掘技术应用于焊接 CAPP 系统中的思想，对 CAPP 系统数据依据深层次的工艺主题进行系统的组织，从焊接数据库中挖掘出隐含的、有用的、尚未发现的信息和知识，提取与焊接 CAPP 相关的工艺规则以满足工艺设计与管理的需求，为智能

化 CAPP 提供一种新的实现方法。

鉴于以上原因，本文以具有代表性的汽车行业的车身焊接工艺设计活动为背景，应用数据挖掘技术，研究和开发一种更为智能化、集成化、可重构的创成式焊接 CAPP 系统。

1.6.1 课题研究的工作内容

课题通过对数据挖掘技术的深入分析，将数据挖掘应用于焊接 CAPP 数据库中，采用关联规则挖掘方法对汽车车身的焊接工艺数据进行挖掘，发现基本工艺单元，建立以数据挖掘技术为依托的创成式 CAPP 系统，优化焊接工艺流程，以期为提高 CAPP 系统的应用效率提供有力的技术支持。

焊接 CAPP 系统的研究和数据挖掘技术是本课题的研究基础，当然目前数据挖掘技术的发展状况和已经取得的成果也是本课题能够开展的必须的技术条件。

1.6.2 课题研究的关键问题和处理方法

针对焊接工艺数据库数据量大，格式不一的特点，焊接工艺关联规则的数据挖掘方法和算法是研究的关键问题之一；另外，如何建立基于数据挖掘技术的创成式焊接 CAPP 系统也是课题研究的关键。

应用数据挖掘技术来处理焊接工艺数据是一个比较复杂的过程，首先从焊接工艺数据库中提取部分典型数据，根据焊接数据库及焊接工艺设计的特点，在经典 Apriori 算法基础上研究适合焊接工艺数据挖掘的算法和实现方法。运用 Delphi 数据库开发软件建立焊接 CAPP 系统的开发平台，以汽车车身的典型焊接工艺设计案例作为研究对象，将数据挖掘技术应用于焊接 CAPP 系统中。

1.7 本章小结

本章从 CAPP 系统原理及基本类型出发，介绍了焊接 CAPP 系统在国内外的的发展历程，将计算机辅助工艺设计引入焊接领域，阐述了 CAPP 系统构成以及焊接 CAPP 系统的特点。本章进一步阐述了焊接 CAPP 的发展趋势以及现有焊接 CAPP 系统存的不足之处。面对这些问题，提出了课题的研究内容及意义，以及针对关键问题的处理方法，为将数据挖掘技术引入到焊接 CAPP 系统埋下伏笔。

第二章 数据挖掘原理

CAPP 系统的数据库中存储了大量的工艺设计资料, 如何充分利用这些宝贵的数据和信息是一项非常有意义的研究工作。数据挖掘作为信息领域的一个活跃课题, 能够为决策者提供重要的、有价值的信息或知识, 从而产生不可估量的效益。通过对数据挖掘技术的深入分析, 现提出将先进的数据挖掘技术引入到焊接生产领域, 从大量的生产数据中获得潜在的规律性和发展趋势, 挖掘基本的工艺单元来指导焊接生产。本章详细讨论了关联规则的定义和改进的 Apriori 挖掘算法, 以期为提高焊接 CAPP 系统的应用效率提供有力的技术支持。

2.1 数据挖掘概述及常用方法

数据库中的知识发现 (Knowledge Discovery in Data-base, KDD), 即数据挖掘 (Data Mining, DM), 是近年来数据库系统和智能技术的最前沿研究内容之一。数据挖掘, 是指从数据集中识别出有效的、新颖的、潜在有用的、最终可理解的模式非平凡过程。简单的说, 就是从数据库的大量数据中挖掘和学习有价值的和隐含的知识^[19-20]。数据挖掘过程通常包括以下步骤。

- (1) 数据清理: 消除噪声或不一致数据。
- (2) 数据集成: 将多种数据源组合在一起。
- (3) 数据选择: 从数据库中检索与分析任务相关的数据以减少数据挖掘时的工作量。
- (4) 数据变换: 通过数据汇总等方法将数据变换或统一成适合挖掘的形式。
- (5) 数据挖掘: 使用不同的数据挖掘算法从大量数据中提取用户需要的知识。
- (6) 知识评估和表示: 将挖掘的知识根据一定条件进行筛选, 并以用户可以理解的方式表示。

2.1.1 数据挖掘技术概述

数据挖掘技术按照挖掘的数据库类型、挖掘的知识类型、采用的挖掘技术以及应用场合等有不同的分类。通常按照挖掘的知识类型的不同, 数据挖掘主要分为以下几类^[21]:

(1) 广义知识

广义知识指类别特征的概括性描述知识, 根据数据的微观特性发现其表征的、带有普遍性的、较高层次概念的、中观和宏观的知识, 是对数据的概括、精炼和抽象。

广义知识的发现方法和实现技术有很多, 如数据立方体、面向属性的归纳等。该方法的基本思想是实现某些常用的代价较高的聚类函数的技术, 诸如计数、求和、平均、最大值等, 并将这些实现视图储存在多维数据库中。另一种

广义知识发现方法是加拿大 SimonFraser 大学提出的面向属性的归约方法。这种方法以类 SQL 语言表示数据挖掘查询, 收集数据库中的相关数据集, 然后在相关数据集上应用一系列数据推广计数进行数据推广, 包括属性删除、概念树提升、属性阈值控制、计数及其他聚类函数传播等。

(2) 关联知识

它反映一个事件和其他事件之间依赖或关联的知识。如果两项或多项属性之间存在关联, 那么其中一项的属性值就可以依据其他属性值进行预测。最为著名的关联归总发现方法是 R.Agrawal 提出的 Apriori 算法。数据关联是数据库中存在的一类重要的可被发现的知识。关联分为简单关联(如买啤酒的男人中 95% 的人会买土豆片等佐酒小菜)、时序关联(如 A 股票连续上涨两天且 B 股股票不下跌, 则第三条 C 股票上涨的可能性为 75%)、因果关联。关联分析的目的是找出数据库中隐藏的关联网。有时不知道数据库中的数据是否存在精确的关联函数, 即使知道也是不确定的, 因此关联分析生成的规律带有可信用度。

(3) 分类知识

它反映同类事物共同性质的特征型知识和不同事物之间的差异型特征知识。最为典型的分类方法是机遇决策树的分类方法。它是从实例集中构造决策树, 是一种有指导的学习方法。该方法先根据训练子集(又称为窗口)形成决策树。如果该树不能对所有对象给出正确的分类, 那么选择一些例外加入到窗口中, 重复该过程一直到形成正确的决策集。最终结果是一棵树, 其叶结点是类名, 中间结点是带有分枝的属性, 该分枝对应该属性的某一个可能值。其中最为典型的决策树学习系统是 ID3, 它采用自顶向下不回溯策略, 能保证找到一个简单的树。

(4) 预测型知识

它根据事件序列型数据, 由历史的和当前的数据去推测未来的数据, 也可以认为是以时间为关联属性的关联知识。

时间序列预测方法有经典的统计方法、神经网络和机器学习等。1968 年 Box 和 Jenkins 提出了一套比较完善的时间序列建模理论和分析方法, 这些经典的数据方法通过建立随机模型, 如自回归模型、自回归滑动平均模型、求和自回归滑动平均模型等, 进行时间序列的预测。由于大量的时间序列是非平稳的, 其特征参数和数据分布随着时间的推移而发生变化, 因此, 仅仅通过对某段历史数据的训练, 建立单一的精神网络预测模型, 还无法完成准确的预测任务。为此, 人们提出了机遇统计学和精确性的再训练方法, 当发现现存预测模型不再适用于当前数据时, 对模型重新训练, 获得新的权重参数, 建立新的模型。也有许多系统借助并行算法的计算优势进行时间序列预测。

(5) 偏差型知识

偏差型知识是对差异和极端特例的描述，揭示事物偏离常规的异常现象，如标准类外的特例，数据聚类外的离群值等。所有这些知识都可以在不同的概念层次上被发现，并随着概念层次的提升，从微观到中观、到宏观，以满足不同用户不同层次决策的需要。

2.1.2 数据挖掘常用方法

数据挖掘涉及到很多交叉学科，包括机器学习、数理统计、神经网络、数据库、模式识别、粗糙集、模糊数学等相关技术。针对不同的应用问题要采用不同的数据挖掘方法和算法，主要的数据挖掘方法有以下几种^[22]：

(1) 关联规则挖掘算法

基本的关联规则挖掘算法是 Apriori 算法，它依赖于有效的频繁数据项集的产生，开辟了关联规则数据挖掘的先河。关联规则可以发现可以分为两步：第一步是迭代识别所有的频繁项目集，要求频繁项目集的支持率不低于用户设定的最低值；第二步是从频繁项目集中构造可信度不低于用户设定的最低值的规则。识别或发现所有频繁项目集是关联规则发现算法的核心，也是计算量最大的部分。

(2) 判定树方法

判定树方法就是利用训练集生成一个测试函数，根据不同取值建立树的分枝，在每个分枝集中重复建立下层节点和分枝，这样生成一颗判定树。然后对判定树进行剪枝处理，最好把判定树转化为规则，利用这些规则可以对新实例进行分类。国际上最有影响的判定树方法是 ID3 算法。

(3) 人工神经网络方法

人工神经网络模仿生物神经元结构，以数据挖掘模型和 Hebb 学习规则为基础，建立前馈式网络、反馈式网络、自组织网络三大类多种神经网络模型。人工神经网络的知识体现在网络连接的权值上，是一个分布式矩阵结构；人工神经网络的学习体现在神经网络权值的逐步计算上（包括反复迭代或累加计算）。在数据挖掘中可用来进行分类、聚类、特征采掘等。

(4) 遗传算法

遗传算法是一种优化技术，它利用生物进化的一系列概念进行问题的搜索，最终达到优化的目的。遗传算法由三个基本算子构成。即繁殖（选择）算子、交叉（重组）算子、变异（突变）算子。它已在优化算法、分类机器学习上发挥了重要的作用。

(5) 粗集方法

粗集理论是一种处理模糊和不定向问题的新型教学工具。在数据库中，把行元素看成对象（分为条件属性和决策属性）。等价关系 R 定义为不同对象在某个（或几个）属性上取值相同，这些满足等价关系的对象组成的集合称为该

等价关系 R 的等价类。条件属性上的等价类 E 与决策属性上的等价类 Y 之间存在三种情况，即下近似，意味着 Y 包含 E ；上近似，意味着 Y 和 E 交非空；无关，意味着 Y 和 E 交空。下近似建立确定性规则，上近似确立不确定性规则，无关不存在规则。利用粗集理论可以处理数据的简化、相关、评价、近似、差异等问题。

(6) 可视化方法

可视化就是采用图形方式将信息模式、数据的关联或趋势呈现给决策者，决策者通过可视化技术交互地分析数据关系。这对揭示数据的具体情况、数据的内在本质及规律起到重大的作用。

2.2 关联规则 Apriori 算法原理

在交易数据库 D 中，假设含有 n 次交易（即 n 个记录），且每一个交易都有一个惟一的交易 ID 号，称为 TID（即主键）。每次交易记为 T ，均为一些商品的集合，每个商品称为一个数据项（Item）。设 $I=\{i_1, i_2, \dots, i_m\}$ 是所有数据项的集合，定义数据项集（Itemset）是数据项的非空集合，数据项集中数据项的数目称为该数据项集的长度。如果数据项集中含有 m 个数据项，则称 X 为 m 数据项集或称 m 长度数据项集。显然 D 中的每一次交易都对应一个数据项集，为讨论问题方便我们假设 T 中的数据项不重复且为有序排列^[23]。

定义 1: $I=\{i_1, i_2, \dots, i_m\}$ 称为数据项集（Itemset），其中 i 为数据项，包含 k 个项的项集称为 k -项集；

定义 2: $D=\{t_1, t_2, \dots, t_m\}$ 称为事物集，其中每一个事物 t_i 对应一个数据项集，即 $t_i \in I$ ，每个 t_i 都有一个唯一的标识符 TID 与其对应。项集的出现频率是包含项集的事务数，简称为项集的频率或支持计数；

定义 3: 关联规则是形如 $A \Rightarrow B$ 的蕴含式，其中 $A \subset I, B \subset I$ ，并且 $A \cap B = \emptyset$ ；

定义 4: 规则 $A \Rightarrow B$ 在事务集 D 中成立的支持度 $\text{support} = P(A \cup B)$ ，即 D 中包含 $A \cup B$ 的百分比。支持度是对关联规则重要性（或适用范围）的衡量；

定义 5: 规则 $A \Rightarrow B$ 在事物集 D 中成立的置信度 $\text{confidence} = P(B|A)$ ，即 D 中包含 A 事务同时也包含 B 事务的百分比。置信度是对关联规则准确性的衡量。

定义 6: 如果项集满足最小支持度，则称它为频繁项集（Frequent itemset），频繁 k 项集的集合通常记作 L_k ；

定义 7: 同时满足最小支持度（ min_sup ）和最小置信度（ min_conf ）的关联规则称为强规则。

给定一个事务集 D ，关联规则挖掘的任务就是发现支持度和置信度分别大于用户给定的最小支持度（ min_sup ）和最小置信度（ min_conf ）的强关联规则。

2.2.1 经典 Apriori 关联规则算法基本思路

Apriori 算法是一种应用最广泛的关联规则挖掘算法，其核心方法是基于频繁集理论的递推方法，即 k -项集用于搜索 $(k+1)$ -项集，同时为了压缩搜索空间，Apriori 算法基于了一个重要的性质：频繁项集的所有非空子集都必须是频繁的。

在进行关联规则挖掘时，首先找出频繁 1-项集的集合，该集合记作 L_1 ， L_1 用于找频繁 2-项集的集合 L_2 ，而用 L_2 于找 L_3 ，如此下去，直到不能找到频繁 k -项集。该算法需对数据库作多次遍历，每次遍历均包括两个步骤^[24]。

(1) 连接步：为找 L_k ，通过 L_{k-1} 与自己连接产生候选 k -项集的集合。该候选项集的集合记作 C_k 。设 l_1 和 l_2 是 L_{k-1} 中的项集，如果

$(l_1[1]=l_2[1]) \wedge (l_1[2]=l_2[2]) \wedge \dots \wedge (l_1[k-2]=l_2[k-2]) \wedge (l_1[k-1] < l_2[k-1])$ ， L_{k-1} 的元素 l_1 和 l_2 就是可连接的。条件 $(l_1[k-1] < l_2[k-1])$ 是为了保证不产生重复。连接 l_1 和 l_2 产生的结果项集是 $l_1[1]、l_1[2] \dots l_1[k-1]、l_2[k-1]$ 。

(2) 剪枝步： C_k 是 L_k 的超集，即它的成员可以是也可以不是频繁的，但所有的频繁 k -项集都包含在 C_k 中。遍历数据库，确定 C_k 中每个候选的计数，从而确定 L_k （即根据定义，计数值不小于最小支持度计数的所有候选是频繁的，从而属于 L_k ）。如果 C_k 很大，为了压缩 C_k ，就必须使用 Apriori 性质。因此，如果一个候选 k -项集的 $(k-1)$ -子集不在 L_{k-1} 中，则该候选也不可能是频繁的，从而可以由 C_k 中删除。

2.2.2 Apriori 算法描述

输入：事物数据库 D ；最小支持度阈值 min_sup 。

输出： D 中的频繁项集 L (L_k 表示： K -频繁项目集； C_k 表示： k -项目集)

[25]

```
(1)  $L_1 = \text{find\_frequent\_1-itemsets}(D)$ ; //产生频繁 1-项集
(2) for ( $k=2$ ;  $L_{k-1} \neq \emptyset$ ;  $k++$ )
    {
(3)  $C_k = \text{apriori\_gen}(L_{k-1}, \text{min\_sup})$ 
For each itemset  $I_1 \in L_{k-1}$ 
For each itemset  $I_2 \in L_{k-1}$ 
    If  $(I_1[1]=I_2[1]) \wedge (I_1[2]=I_2[2]) \wedge \dots \wedge (I_1[k-2]=I_2[k-2]) \wedge (I_1[k-1] < I_2[k-1])$  then
    {
         $C = I_1 \cup I_2$  //c 是  $C_k$  中的一条记录
    If has_infrequent_subset(c,  $L_{k-1}$ ) then
//候选项集中如果含有非频繁项集，则候选项集也是非频繁的
Delete c;
Else add c to  $C_k$ ;
    }
Return  $C_k$ ;
```

```

(4) for each transation  $t \in D$ //扫描 D, 对每个事务 t 操作
    {
(5)  $c_t = \text{subset}(C_k, t)$ ;//subset 函数找出事务中是候选的所有子集
(6) for each candidate  $c \in C_t$ 
(7)  $c.\text{count}++$ ;//对每个候选项集计数
    }
(8)  $L_k = \{c \in C_k, c.\text{count} \geq \text{min\_sup}\}$ 
    }
(9) return  $L = L \cup L_k$ 

```

补充: procedure has_infrequent_subset(c, L_{k-1})

```

(1) for each (k-1)-subset s of c//求(k-1)-项目集的所有(k-2)-项子集
(2) if  $s \neq L_{k-1}$  then
(3) return true;
(4) return false

```

2.2.3 Apriori 算法的缺点

Apriori 算法虽然使用了“一个项集是频繁项集当且仅当它的所有子集都是频繁项集”的性质来优化,但它还是存在一些缺陷^[26]:

(1) 可能产生大量的候选集

当长度为 1 的频繁项集有 1000 个的时候,长度为 2 的候选集个数会超过 10M;如果需要产生一个很长的规则的时候,要产生的中间元素也是巨大量的。

(2) 无法对稀有信息进行分析

由于使用了参数 min_support,所以就无法对小于 min_support 的事件分析;而如果将 min_support 设成一个很低的值,那么算法的效率就成了一个很难处理的问题。

(3) 需要多次扫描数据库

验证候选集是频繁项集的过程是算法的一个瓶颈,需要多次扫描可能很大的交易数据库。如果最多包含 10 个项,那么就需要扫描交易数据库 10 遍,这需要很大的负载。

(4) 支持一信任框架理论

通过支持度和信任度来挖掘关联规则的方法,就是我们通常所说的“支持一信任框架理论”,基于这个理论挖掘出来的规则具有一定的欺骗性,它只是给定了条件概率的估计,而并不估量前件与后件之间蕴涵的实际强度,利用这个理论挖掘出的规则具有许多局限性,还存在规则冗余的问题。

为了克服这些缺点,研究人员提出了很多优化方法,例如基于划分的方法,基于哈希表的方法,基于采样的方法,基于缩减数据库的方法,引入提升度指标剔除错误的规则等等。

2.2.4 Apriori 改进算法描述

本文提出的 Apriori 改进算法，将 Apriori 算法中对事务数据库扫描和模式匹配计算候选集的支持度，改进为利用 SQL 语句直接对数据库中的数据进行操作，可以省去把数据从数据库中提取再用程序控制实现的过程。SQL 语言只利用其提供的 9 个动词就可以满足用户从建库到查询、数据维护和简单统计等对数据库的操作需求，无须再用烦琐的程序来控制循环。对关系表的查询来计算候选集的支持度，将频繁 n -项集及其支持度保存在关系表中，在生成关联规则时判断可信度是否大于设置的最小可信度的阈值。

引入 SQL 语句仅扫描一次数据库，并删除冗余规则产生强关联规则的改进 Apriori 算法描述：

$S[i]$: i 维数据源，由中间集 $M[i-1]$ 与 $M[1]$ 连接生成；

$R[i]$: i 维结果集，扫描数据源 $S[i]$ ，提取大于支持度阈值的项集组成；

$M[i]$: i 维中间集，保留数据源 $S[i]$ 中存在于结果集 $R[i]$ 中的项集；

Support: 支持度，描述项集 A 在所有的事务中出现的概率；

Confidence: 置信度，描述在出现了项集 B 的事务 A 中，项集 $(A-B)$ 同时出现的概率有多大；

(1) 输入：数据源， $S[1]$ ；最小支持度阈值 \min_sup 。

(2) For($i=1$; $S[i] \neq \emptyset$; $i++$)

{

(3) $R[i] = \text{Generate_Resultsets}(S[i], i)$;

 Procedure Generate_Resultsets($S[i], i$)

 Insert into $R[i]$

 Select $item_1, item_2, \dots, item_i, count(*)$ from $S[i]$

 Group by $item_1, item_2, \dots, item_i$

 Having $count(*) \geq \min_sup$

 Order by $item_1, item_2, \dots, item_i$;

(4) $M[i] = \text{Generate_Middlesets}(S[i], R[i], i)$;

 Insert into $M[i]$

 Select $S[i].id, S[i].item_1, \dots, S[i].item_i$ from $S[i], R[i]$

 Where $S[i].item_1=R[i].item_1$ and $\dots, S[i].item_i=R[i].item_i$

 Order by $S[i].id$;

(5) $S[i+1] = \text{Generate_Sourcesets}(M[i], M[1], i+1)$;

 Insert into $S[i+1]$

 Select $M[i].id, M[i].item_1, \dots, M[i].item_i, M[1].item_1$ from $M[i], M[1]$

 Where $M[i].id = M[1].id$ and $M[i].item_i < M[1].item_1$

 Order by $M[i].id$;

 If $M[i] = \emptyset$ or $S[i+1] = \emptyset$

 then break;

}

(6) For($i=2$; $R[i] \neq \emptyset$; $i++$)

 For $R[i]$ 中的每个项集 A

 For A 中的每个非空真子集 B


```

{ Support =  $\frac{A.count}{n}$ 
  Confidence =  $\frac{A.count}{B.count}$ 
  If confidence  $\geq$  min_conf then
    return: R[i]的各条记录:
  Else
    Return :false;
}

```

补充: procedure has_infrequent_subset(c, Lk-1)

- (1) for each (k-1)-subset s of c //求(k-1)-项目集的所有(k-2)-项子集
- (2) if s \neq Lk-1 then
- (3) return true;
- (4) return false

步骤(2)是输入原始数据源，也就是一维数据源 S[1]；步骤(2)使用函数 Generate_Resultsets 产生一维结果集 R[1]，通过利用 item1 对 S[1]进行分组查询得到；步骤(4)使用函数 Generate_Middlesets 从 i 维数据源 S[i]中筛选出含有结果集 R[i]的那一部分数据源，筛选的方法是利用 S[i]和 R[i]的对应项相等；步骤(5)使用函数 Generate_Sourcesets 将 i 维中间集 M[i]和原先的一维中间集 M[1]组合生成 i+1 维数据源 S[i+1]，方法是使用 M[i]和 M[1]的 id 相等，而且 M[1]的项目不会和 M[i]的项目重复。算法改进了 Apriori 算法每次都扫描原始的数据集，而基于 SQL 的 Apriori 算法除了第一次是扫描原始数据集，而其余执行都是扫描中间集 M[i]的到的数据源。由于 M[i]是大于最小支持度的频繁项目集，根据大项目集的基本原理，M[i]可以代替原始的数据集，有效缩小搜索空间，从而提高算法的效率。

2.3 决策树算法原理

决策树方法最早产生于 20 世纪 60 年代，到 70 年代末，由 J.Ross.Quinlan 提出了 ID3 算法，算法的目的在于减少树的深度，但是忽略了叶子数目的研究。C4.5 算法在 ID3 算法的基础上进行了改进，对于预测变量的缺值处理，剪枝技术、派生规则等方面作了较大改进，既适合于分类问题又适合于回归问题。

2.3.1 决策树的构造方法

决策树构造的输入是一组带有类别标记的例子，构造的结果是一棵二叉树或多叉树。二叉树的内部节点（非叶子节点）一般表示为一个逻辑判断，如形式为 (a=aj) 的逻辑判断，其中 a 是属性，aj 是该属性的所有取值；树的边是逻辑判断的分支结果。多叉树 (ID3) 的内部结点是属性，边是该属性的所有取值，有几个属性值就有几条边。树的叶子节点都是类别标记。

由于数据表示不当、有噪声或者由于决策树生成时产生重复的子树等原因，

都会造成产生的决策树过大。因此，简化决策树是一个不可缺少的环节。寻找一棵最优决策树，主要解决以下三个最优化问题：

生成最少数目的叶子节点；

生成的每个叶子节点的深度最小；

生成的决策树叶子节点最少且每个叶子节点的深度最小。

ID3 算法是最为经典的决策树学习系统，采用自上而下的不回溯策略，能保证找到一颗简单的树。其工作过程为：首先选出最有判别力的测试属性，把数据分成多个子集，每个子集又选择最有判别力的另一测试属性进行划分，如此一直进行到最后划分出的所有子集仅包括同一类型的数据为止，最后得到一棵决策树，可以用它对新的例子进行分类^[27]。ID3 中测试属性的选取是以信息熵的下降速度为标准。信息熵的下降也就是信息不确定性的下降。

ID3 是在两个假设的基础上采用基于信息熵的方法，把对象集 C 分为两类，其中 P 类对象有 p 个，N 类对象有 n 个^[28]，假设为：

任何一个正确的决策树，对所有对象依据其所占的比例来进行分类，即对于任意的对象属于 P 类的概率是 $\frac{p}{p+n}$ ，属于 N 类的概率是 $\frac{n}{p+n}$ 。

当用一个决策树对一个对象进行划分时，返回一个类别。一个决策树可以被看成是分类信息的来源，决策树对做出正确划分需要的期望信息熵是：

$$I(p, n) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \quad (1)$$

假如条件属性 A 有属性值 $\{A_1, A_2, A_3, \dots, A_v\}$ ，被用做决策树的根时，它将对象子集划分为 $\{C_1, C_2, C_2, \dots, C_v\}$ ，其中， C_i 是属性 A 取 A_i 的对象，包括 P 类 p_i 个对象，包括 N 类 n_i 个对象，那么子树 C_i 的期望信息熵是 $I\{p_i, n_i\}$ ，对于树根 A 划分整个树所需的期望信息熵

$$E(A) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} I(p_i + n_i) \quad (2)$$

通过属性 A 来分支的信息增益是

$$\text{Gain}(A) = I(p, n) \quad (3)$$

$E(A)$ 选取属性作为决策树节点的原则就是能够得到最大信息增益的属性来分类形成树，然后 ID3 算法依次检测每个候选属性，用相同的处理过程递归地形成决策子树 $C_1, C_2, C_3, \dots, C_v$ 。

2.3.2 决策树的改进

本文中采用决策树 C4.5 算法进行知识的发现，C4.5 算法继承了 ID3 算法的全部优点，并且进行了改进，提出了用增益比例代替增益进行属性的选择，增益比例的定义为：

$$\text{GainRatio} = \frac{\text{Gain}(X, a)}{\text{SplitInfo}(X, a)} \quad (4)$$

其中确定属性 A 本身需要的信息熵是

$$\text{SplitInfo}(A) = - \sum_{i=1}^v \frac{p_i + n_i}{p + n} \log_2 \frac{p_i + n_i}{p + n} \quad (5)$$

采用增益比去划分属性得到的决策树，其中每个节点取具有最大信息增益比的属性。

2.3.3 决策树的修剪

完整的决策树生成以后，一般不能用于对新数据分类或预测。主要原因是完整的决策树对训练样本特征的过度拟合。对决策树进行修剪也是最常用的简化决策树的方法。主要包括预先剪枝和后剪枝。预先剪枝是在建树的过程中，当信息增益达到某个预先设定的阈值时，节点不再继续分裂，内部节点成为一个叶子节点。后剪枝是在允许决策树得到最充分生长的基础上，根据一定的规则，剪去决策树中的那些不具有一般代表性的叶节点或分枝^[29]。

遍历形成的决策树，从根到叶就发现若干条路径，每一条路径对应一条规则，整棵树就形成了一组析取表达式规则，然后详审规则集去发现最有用的子集，最后的规则集用计算机可读的格式存储在一个文件中。

2.4 本章小结

本章主要介绍了数据挖掘理论的基本技术和常用方法，重点介绍了关联规则算法和决策树算法的基本原理和算法缺点，并在此基础上提出基于 Apriori 算法的改进算法，将 Apriori 算法中对事务数据库扫描和模式匹配计算候选集的支持度，改进为利用 SQL 语句直接对数据库中的数据进行操作，省去把数据从数据库中提取再用程序控制实现的过程。为了读者更好的理解系统数据挖掘模块的设计思想，系统还详细介绍了改进算法的思路及语言描述。

第三章 基于数据挖掘的焊接 CAPP 系统

汽车工业作为当今世界经济支柱性产业之一，在发达国家国民经济中占有举足轻重的地位，而在中国，汽车工业现代化的发展起步于上世纪 50 年代，初期发展缓慢，到 80 年代后期方初具规模。到 90 年代初，中国已经能生产各种类型汽车，包括重、中、轻、微型货车、客车、轿车和专用车。1995 年，总产量达到 140 万辆，汽车生产厂家达到 122 家，其中前 10 名厂家产量占总计的 74%，生产成高度集中之势。中国政府于 1994 年 2 月颁发了《汽车工业政策》，其目标是到 2010 年使其成为国民经济的支柱产业，并带动其他相关产业的发展。由此可以预料，在政策倾斜上，汽车工业拥有相当大的发展前景。

尽管国内汽车生产制造能力越来越强，但是国内汽车焊接水平与国外相比仍存在一定的差距，这个差距不仅仅来自于焊接技术的落后，也在于国内汽车产业的创新意识仍远远落后于发达国际，在汽车自动化焊接生产线方面的自主开发还属于起步阶段，大部分汽车生产厂商在这方面都是引进国外技术和产品，而随着汽车工业的发展，汽车焊装生产线必然会逐渐向着全自动化方向发展，所以目前越来越多的国内汽车制造商开始尝试用自己的技术来开发焊接生产线。

作为一项难以精确量化的热加工技术，焊接工艺设计必须依靠积累的经验 and 知识，配以必要的理论来解决。焊接工艺设计涉及的技术环节较多，对于母材、接头形式、热处理等数值因素和标准、规范、专家经验等非数值因素，以及其他一些相关的模糊因素都要综合考虑。因此，将优化设计方法和数据挖掘技术引入焊接工艺设计中，充分发挥计算机和人的各自专长，可以帮助工艺人员选定适用的焊接工艺，局部摆脱或完全摆脱对经验的依赖，进行焊接工艺的规范设计^[30]。

针对设计与管理而开发的焊接 CAPP 系统，是一个直接面向生产实际和工程设计人员的应用系统。本系统是根据汽车焊接工艺设计的特点和 CAPP 的研究现状，在零件信息、工艺信息、工艺知识的基础上运用数据挖掘技术建立的以汽车为应用背景的创成式焊接 CAPP 系统。

3.1 创成式焊接 CAPP 系统原理

3.1.1 创成式 CAPP 系统

创成式（也称生成式）计算机辅助工艺设计系统是依靠输入零件图形及加工要求等信息，通过预先存入计算机系统数据库或知识库中的一系列与设计零件工艺有关的工艺决策规则，诸如确定母材类型、设备、方法等，自动地为零件制订出工艺过程的系统。与此同时，根据具体零件，系统能自动产生零件所需要的各个工序和加工工序，自动提取制造知识，自动完成加工方法选择（包括机床选择、焊接方法以及其它方法等）、工具选择和加工过程的最优化；

通过应用决策逻辑，可以模拟工艺设计人员的决策过程^[31]。

创成式 CAPP 系统的工艺设计过程是整个系统的核心部分，工艺设计过程分为零件信息输入、工艺过程创成、修改工艺规程和输出工艺规程几个阶段。其流程图如图 3-1 所示：

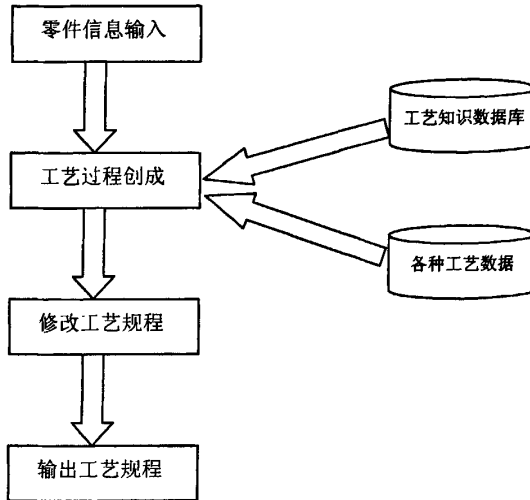


图3-1 创成式CAPP系统原理图

3.1.2 基于数据挖掘技术的创成式焊接 CAPP 系统

本文设计的焊接 CAPP 系统是一个将 CAPP 与数据挖掘技术紧密集成的一个综合系统，它利用互联网以及大量离线数据为依托，利用数据挖掘技术，综合加工信息、挖掘海量信息中隐含的可靠的，有用的工艺信息，自动地为一个新零件制定出工艺过程的系统。系统可以根据具体的零件，自动产生零件所需要的加工工序，提取制造知识，完成焊接方法的选择，工艺参数的设计，以及工艺流程的优化等。通过应用决策逻辑，可以模拟工艺设计人员的决策过程。

基于数据挖掘技术的创成式焊接 CAPP 系统在原有创成式 CAPP 系统基础上增加了数据挖掘模块，如图 3-2 所示：

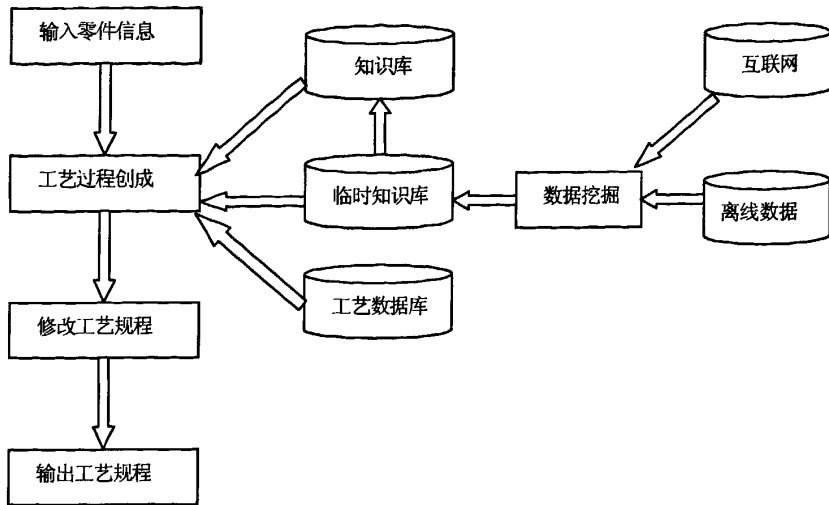


图3-2 基于数据挖掘技术的创成式CAPP系统原理图

(1) 零件信息输入 在对零件进行工艺设计之前，设计人员要先确定该零件的一些信息，包括母材的种类，成分，组织性能，力学性能，焊材的成分与性能，保护气体特性用途等。系统再根据零件信息的描述和决策逻辑与方法的不同，对每一类具体零件进行工艺决策系统的开发，从而使生成的工艺信息更准确可行。

(2) 数据挖掘技术 由于英特网上资源的共享及工厂多年的生产实践经验，大量的数据信息可以通过在线查找和文献查阅的方式获得，而传统的经验无法分析出隐含在数据中的工艺规则，如何利用这些丰富的数据资源，是数据挖掘技术应用于焊接生产过程的目的。关联规则改进算法，是基于工艺优化进行设计的。例如，流水线式的生产管理，使得部分零件可以通过一道道工序自动加工生产而成，工人可以从嘈杂的生产环境中解脱出来，而零件在流水线生产过程中，一道道工序的设计，工人工位的布置等，都是从大量的生产实践中获得的。对于新零件的设计，如果只凭借工艺人员的传统经验来进行工艺参数的设置，肯定会造成设计周期的延长和生产资源的浪费。通过数据挖掘技术，提取已有的相似零件信息，利用算法逻辑可以挖掘出每一道工序或者参数之间的潜在规则，对新零件的设计，可以提出一套比较科学的工艺设计方案进行参考。

(3) 工艺过程创成 工艺过程创成作为创成式焊接 CAPP 系统的核心环节，其创成过程及方法是尤为重要的。确定零件种类后，就可以进行工艺路线的设计。系统的零件信息取自各种工艺数据库，例如零件的特征信息，如坡口参数以及接头类型等，这些数据是被存放在数据库中的。本系统在原有创成式 CAPP 系统工艺数据库的基础上增加了知识库和临时知识库的部分，一方面将数据挖掘的关联规则存储在临时知识库中，为工艺过程创成提供参照；另一方面，临时知识库还可以不断扩充知识库，为今后的工艺设计提供参考。系统对零件的加工路线或工艺参数进行调整后，可以人工参与修改，最终可以得到比较满意

的工艺路线。

(4) 修改工艺规程 工艺过程生成经过以上步骤后可生成一棵能够表示该零件所有特征加工先后顺序的树，修改确定后生成工艺规程主干，用户如不满意可对其进行添加、删除和修改等操作，并对该工艺规程主干进行扩充，直到满意为止。

(5) 输出工艺规程 在完成了全部的工艺设计任务以后，系统按照规定的格式打印。

3.2 系统功能分析及总体结构设计

焊接工艺主要是指焊接的先后次序和相应的内容，也就是焊接件的实际生产过程，通常不包括零部件的具体加工，如下料、成型等。焊接工艺的设计和管理是一个相当复杂的过程，目标系统中涉及的数据很多，在设计过程中，要确定具体的工艺路线和内容，包括焊接母材，母材种类，焊接材料的种类、牌号及规格，焊接方法，接头类型，坡口形式，焊接工艺参数，焊前及焊后热处理，从而保证整体焊接工艺的完整性和有效性。因此，焊接工艺的设计和管理所涉及的信息量是相当大的。焊接工艺设计与一般的机械加工方法相比，具有以下几个特征^[32]：

(1) 焊接结构件形状复杂、特征变化很多，结构特征的表述难度大。焊接结构件的制造是装配和焊接的组合加工过程，它们即有一定的独立性，又有极为密切的关系。相同的结构采用不同的焊接顺序，也会有不同的参数，这些都增加了工艺设计的难度。

(2) 焊接材料种类繁多，环境条件复杂，而且它们的变化对焊接参数、焊接工艺流程的影响很大。由于焊接属于热加工，是一个经验性很强的工作，难以找到实用、可行的规律，也很难抽象出一个数学模型进行仿真，加工顺序和工艺参数难以确定。

(3) 焊接结构件的制造，与焊接相关联的辅助工序，如组装定位、焊前预热、焊后热处理等，都具有一定的随机性。然而，焊接结构件本身具有相对的稳定性，在同一类焊接结构件中，无论结构还是功能，都具有一定的相似性和继承性，根据结构、功能、工艺相似的原则，同一类焊接结构件的焊接工艺也具有一定的相似性和继承性。

鉴于焊接工艺设计的特征，如何合理有效地收集、管理和利用焊接工艺信息，是本系统的一个重要任务。作为焊接工艺数据库管理系统的重要组成部分，系统功能和结构的设计，不仅对于实现焊接工艺文件的计算机化管理有重要意义，同时也能为数据挖掘模块帮助焊接技术人员正确而迅速地制定焊接工艺提供有力支持。

3.2.1 系统设计思想

基于企业现状和焊接工艺设计的一般特点，针对当前计算机辅助焊接工艺设计系统的不足，吸取它们的优点，确定了如下设计思想：

(1) “面向企业”的思想

此系统是面向企业的设计工作，以具体企业为研究对象开发创成式焊接CAPP系统，并以企业的实际使用效果为目标。开发的CAPP系统所输出的工艺信息应符合工厂焊接生产的习惯和经验，采用企业工艺设计人员和操作人员习惯和熟悉的工艺术语。同时系统应简单实用，以利于用户操作和培训。

(2) “人机协同”的思想

焊接工艺设计涉及的工艺规程比较复杂、知识获取难度较大，经验性也非常强。如果单纯依靠计算机进行工艺设计，生成的工艺可能不尽合理，系统开发的成本也可能比较高。采用“人机协同”的思想，工艺设计人员在系统设计阶段可交互修改工艺信息，实现人的主动性和创造性工作；而数据处理、规则化知识挖掘工作则由计算机来完成。“人机协同”工作，使系统的工作设计更容易，提高系统的适应能力^[33]。

(3) “数据挖掘”的思想

工艺设计是一门经验性强、涉及面广的综合学科，属于典型的多因素复杂问题，目前还不能建立实用的数学模型和相应的通用算法来解决，许多决策依赖于经验知识。在开发焊接工艺设计系统的过程中，应该充分利用企业已有的大量的工艺信息以及技术人员的经验，采用数据挖掘的技术，提取大量潜在的工艺规则，用于指导新工艺的设计过程，提高系统的智能性和实用性。

(4) “快速原型”的思想

快速原型是在初步明确设计需求后，把系统主要功能接口作为设计依据，快速开发出软件初始模型，以后数据库和测试用例对原型系统进行评价、修改、扩充和调整，不断完善原型，扩充系统功能。

3.2.2 系统功能分析

开发以汽车焊接工艺设计为应用背景的创成式焊接CAPP系统，将主要从企业的实际生产过程和技术入手，分析确定焊接CAPP系统的功能需求，进而设计出系统的功能模型和总体结构。经分析，焊接企业对本系统存在如下功能需求：

(1) 能够对焊接工艺文件进行有效的管理。系统对这些焊接工艺文件提供了比较完备的管理手段，以便提高用户对焊接工艺文件的管理水平和效率。它不仅可以对工艺信息进行检索、查阅、输入、存储及删除、修改等；还可对母材型号、化学成分、力学性能、焊材类型、坡口形式及厚度等重要条件进行组合查询。它为CAPP系统组织起完整的工艺设计信息模型，是进行计算机辅助工艺设计的基础。

(2) 进行焊接坡口图形的查询。焊接坡口形式是焊接工艺设计的一项重要因素，坡口形式选择与设计的好坏，直接影响到焊接结构的生产加工工艺、效率及成本。系统提供了对于坡口图形的维护与管理，用户可以添加、删除、更新坡口图形，更改坡口的工艺参数。

(3) 能够模拟焊接技术人员进行焊接工艺的制定。工艺设计是焊接 CAPP 系统的核心部分，主要由它来完成焊接工艺规程的设计。系统根据互联网和工艺知识库中的数据和信息，利用数据挖掘技术分析推理，通过检索、查询或挖掘得到焊接工艺规程。用户可以根据系统提示选择母材类型、母材牌号、焊接位置等初始条件，在与用户不断的交互过程中实现焊接工艺的制定。

(4) 能够进行焊接资料库的维护和管理。焊接工艺制定过程中，需要了解一定的信息，如母材牌号、化分成分及机械性能，焊材牌号、化学成分、主要用途，焊接方法、焊接设备等等。

(5) 能够为用户提供简单实用的系统帮助功能。此部分属于系统的辅助功能模块。

3.2.3 系统结构设计

焊接CAPP系统其实质也是一组计算机程序，所以应当而且必须遵循软件开发的一般规律，对于本系统来说亦是如此。基于数据挖掘的焊接CAPP系统是一个涉及焊接资料查询、焊接工艺设计、焊接规则提取等方面的复杂系统。按其内容主要有以下几个模块组成：

(1) 数据库系统查询模块

该模块主要是系统运行时所需要的各种焊接资料数据库，包括母材库、焊接材料库、工艺参数库、焊接设备库、焊接方法库等几个综合数据库，还包含一些其它细节的数据库。数据库系统提供一定程度的开放，可满足数据库的扩充。

(2) 焊接坡口工艺设计模块

焊接坡口工艺设计模块是焊接CAPP系统的核心部分，主要由它来完成坡口工艺参数的设计。本模块不仅可以查询坡口的工艺参数，而且可以根据用户输入的初始条件，通过检索或者推理使用户在与系统交互的过程中完成坡口工艺参数的设计工作。

(3) 工艺规则获取模块

作为一个智能数据库系统，本系统的智能体现在运用数据挖掘技术来实现焊接工艺流程的规划，潜在工艺规则获取是本系统的重点部分。在制定焊接工艺时，数据挖掘算法将根据用户输入的工艺信息和数据进行推理，确定这些设计条件参量的合理取值，在与用户交互过程中最终实现焊接工艺的制定。

(4) 系统帮助模块

系统帮助模块是系统的辅助功能模块，阐述了系统的结构、具备的知识以及能够完成的功能，为用户提供了简单实用的系统帮助功能。

3.3 焊接工艺数据库设计

数据库是一组排列成易于处理和读取的相关信息的集合，是计算机应用技术中的一个重要组成部分。对于大量的数据，使用数据库来储存、管理，将比通过文件来储存管理有更高的效率。数据库是整个系统的基础，其设计是否合理是建立数据库管理系统的基础，焊接工艺设计涉及到大量的数据，因此数据库的设计与实现是本系统软件开发的技术关键。

数据库设计的基本任务是：根据企业的信息需求、处理需求和数据库的支撑环境（包括 DBMS、操作系统和硬件），设计出数据模式以及典型的应用程序。信息需求表示企业所需要的数据及其结构；处理需求表示企业经常需要进行的数据处理。数据库设计的基本过程如图 3-3 所示^[34]。

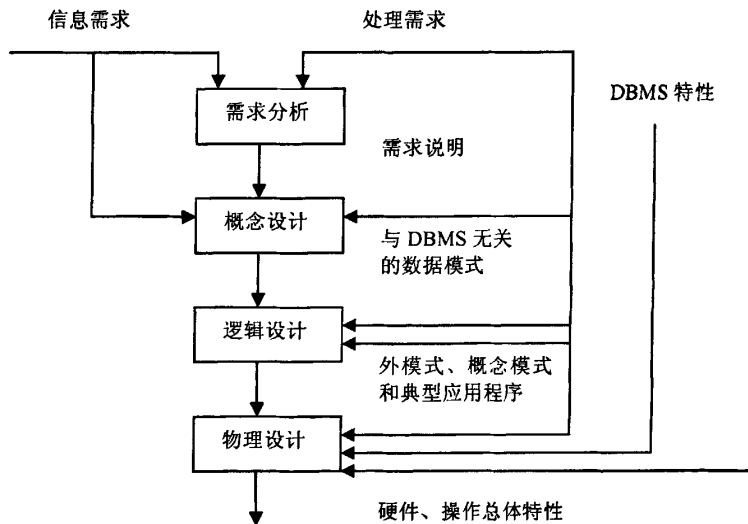


图3-3 数据库设计的基本过程

(1) 需求分析

通过调查，在了解并收集企业信息需求和处理需求的基础上，分析整理成为需求说明。需求说明中包含设计数据库所需的全部原始数据：如数据的名称、标识及描述属性、数据项的类型等。

(2) 概念设计

概念结构设计是整个数据库设计的关键。概念设计的任务是根据需求说明书中采集的素材，使用某种概念数据模型清楚地表示数据之间的相互关系。概念设计通常采用的是E-R方法（Entity-Relationship，即实体—联系），即用E-R图来进行设计。从概念上讲，被分析的对象可看作是由图形表示的一组实体类型和联系类型。在E-R图中，基本的图形元素有3个：实体集合框，联系集合框，

属性集合框。概念结构设计通常分为3个步骤来进行，首先进行实体分析，然后以局部需求为基础来设计局部模式，最后在此基础上生成全局模式，生成总体E-R图并进行优化。

(3) 逻辑设计

根据数据库概念设计的结果，设计出数据库的概念模式和外模式。概念设计与具体的DBMS无关，它不考虑数据与操作之间的关系，仅从概念上明确描述数据之间的关系。而在逻辑设计阶段，首先将用概念数据模型表示的数据及其相互关系转换成DBMS所支持的数据模型的表示方法，然后根据应用中的操作要求及数据的语义约束，对数据的组织结构进行适当调整，包括规范化、逆规范化、关系大小的调整以及某些与现实有关的考虑，形成较为合理的概念模式，最后进行内模式的设计。这个阶段所得到的概念模式和外模式是应用程序访问数据的接口，是整个系统赖以实现的基础，直接影响到数据库的使用性能，而且结构的合理性还直接关系到数据库能否很好地适应应用需求的变化。

(4) 物理设计

数据库物理设计的任务是将由逻辑设计得到的概念模式转换为内模式。数据库的内模式包括记录的存储结构、库文件的存储结构以及访问一个库文件中特定记录的路径。记录的存储结构是指：组成一个记录的数据项如何表示、数据项编码怎样压缩存储，以及记录之间如何连接等。库文件的存储结构是指一个库文件中的记录在存储空间中的位置关系，例如记录是顺序存放还是按照某种算法来定位的，或者将那些经常在同一访问中涉及到的一批记录尽可能地一起存放，形成记录簇集；访问路径是指索引等有效地实现访问特定记录的机制。数据库设计涉及许多方面，从静态的信息需求，到企业的业务政策，各种因素常会产生矛盾，还要考虑到企业未来的发展将会给数据库设计带来的问题。因此数据库设计是一个复杂的、变化的、需要不断改进的创造性过程。

结合系统对数据库的需求分析，在设计过程中遵循以下原则：

(1) 规范性

由于焊接工艺相关信息具有一定的系统性，要求对库中数据的描述做到标准化和规范化，对数据的描述尽可能详尽，数据按统一规范方式输入，计量单位均采用国际标准和公认形式，以保证数据便于统一管理、扩充和交流。

(2) 可扩充性

数据库可方便地进行数据维护、添加、修改和删除。

(3) 查询

不需要具备专门的计算机知识，就能完成对数据库的查询、检索和咨询，使数据库具有优良的实用性。

(4) 选择性

用户能根据自己的设计选择自己需要的焊接工艺文件和焊接工程信息，提

高工作效率和质量。

3.3.1 焊接 CAPP 系统数据库概念结构设计

通过分析制定焊接工艺过程中涉及到的数据和处理过程，以及数据之间的关系，用分类、聚集和概括的抽象机制对需求分析阶段收集到的数据及对数据的处理需求进行分类、组织，得到数据库系统的实体和实体的属性为：

母材：{种类，牌号，化学成分，力学性能，组织性能}

焊条：{种类，牌号，化学成分，力学性能，特性用途}

焊丝：{种类，牌号，化学成分}

焊剂：{种类，牌号，粒度，电流，焊丝}

保护气：{气体名称，成分，性质，适用母材，焊接方法}

焊接坡口：{接头形式，坡口形式，坡口图，工艺参数}

焊接方法：{方法名称，方法定义，工艺参数}

焊接设备：{设备型号，设备名称，设备用途，工作范围}

工艺参数：{焊接电流，焊接电压，焊接速度，气体流量，焊丝伸长度}

焊后热处理：{工艺类型，温度，时间}

分析各个实体之间的关系，确定其各自的联系类型如下：

选定母材后，可以选择不同的焊接方法进行焊接，而一种焊接方法，可以适用于不同的母材，故母材和焊接方法之间是多对多（m：n）的关系。

焊接方法确定后，可以选择不同的焊接材料，而一种焊接材料只能对应一种焊接方法，故焊接方法和焊接填充材料之间是一对多（1：n）的关系。

母材确定后，可以选用不同的焊接接头进行焊接，而一种焊接接头可以适用于不同的母材，故母材和焊接接头之间是多对多（m：n）的关系。

一种焊接方法对应于许多工艺规范参数，而一种规范参数只对应于某一种确定的焊接方法，故焊接方法和焊接工艺规范参数之间是一对多（1：n）的关系。

一种焊接接头可以使用不同的焊接设备，而一种焊接设备适用于多种焊接接头，故焊接接头和焊接设备之间是多对多（m：n）的关系。在对系统需求的详细分析基础上，将以上实体及其属性和实体之间的关系用E-R模型表示，得到如图3-4所示的E-R图，从而建立起数据库的概念模型。

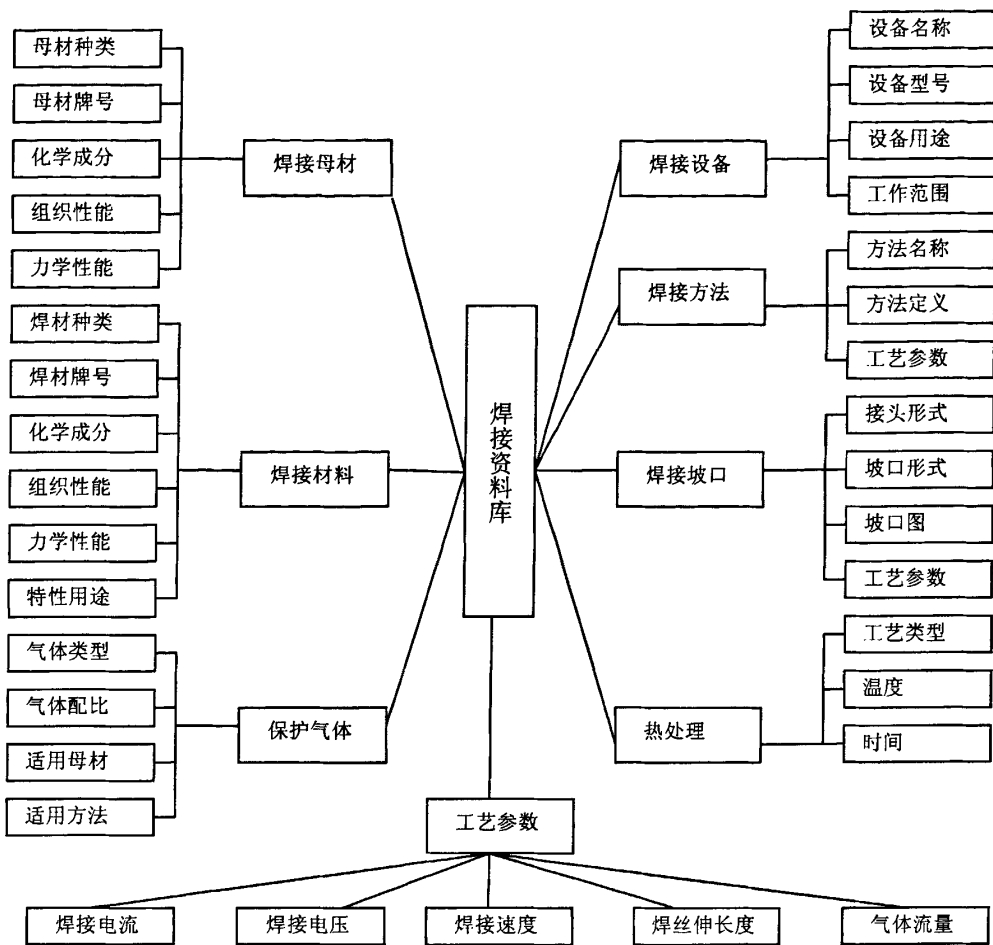


图3-4 焊接工艺数据库E-R模型图

3.3.2 系统数据库逻辑设计与物理设计

逻辑设计的中心任务是找出能表达E-R模型的数据模型，如层次模型、网络模型和关系模型，进行E-R模型向数据模型的转换。即在概念设计的基础上，将概念模型转换为具体DBMS支持的数据类型，包括确定数据项、记录和记录之间的关系等^[35]。本系统采用关系型数据库，故选用关系模型。从E-R模型向关系模型转换的步骤如下。

- (1) 确定各实体的主关键字；
- (2) 写出实体内部属性之间的函数依赖；
- (3) 把函数依赖表达式中的实体名换成相应的主关键字；
- (4) 根据(2)、(3)中各函数依赖关系分组，得到一组关系；
- (5) 完成所有E-R模型向关系的转换后，就形成了关系数据模型。

至于数据库的物理设计，其任务是使数据库的逻辑结构能在实际的物理存

储设备上得以实现，建立一个具有较好性能的物理数据库。数据库物理设计主要解决以下3个问题：恰当地分配存储空间；决定数据的物理表示；确定存储结构。数据库物理设计在很大程度上与选用的数据库管理系统有关。

3.3.3 数据库的建立

数据库的记录包含的信息存放在每个字段中，定义数据库结构设计创建若干字段来存放用户为数据库表中的每一个记录收集的信息，以便有效地对数据库进行管理，主要包括确定数据库字段、定义字段类型及其长度。确定数据库字段时应注意以下几个方面：

(1) 尽可能增大数据库的完备性

数据库系统的目的是创造有用的输出信息，或者是用户能用于总结和分析的信息。所以在建立一个数据库之前，先对每一张用户认为可能有用的报表进行仔细考虑，并考虑进行数据处理时涉及的数据项。因此数据字段应包括：由用户明确提出的项、在任何报表中包括的或任何计算机中使用的项、为一个报表或处理操作确定顺序的项及有助于记录的项。

(2) 数据资源尽可能适应共享的需要

一个数据库在大多情况下不是孤立存在的。可能会调用其它的数据资源，也可能被其它的数据库调用。因此，在建立数据库时，应有利于数据共享。

(3) 尽可能减少数据的冗余度

当字段很多时，同时可建立多个数据库，按共同的字段将它们连接起来，去掉每个数据库在别的文件中可以查到的信息项以避免冗余，同时，数据结构不应该包含重复的多余字段。为此，有条理地处理从用户那里收集来的样本输入、输出格式及任何书面描述，以便建立合适的数据库结构。

(4) 字段命名简单明了

在数据库建立之前，应为字段提出一套命名规则。这些指导规则应尽可能简单，使其他可能要修改系统的程序员容易记住这些规则，同时命名要明确，能概况其含义，避免不必要的令人费解的缩写。

本系统所采用的Database Desktop是Delphi提供的一个功能强大的数据库维护软件，具有丰富的数据类型，可以用来生成、修改和查询本地数据表，创建数据库别名，进行数据操作（如编辑和删除数据等，）以及修改数据表的结构等。另外，Database Desktop提供的QBE（query by example）可以让不熟悉SQL语句的用户建立自己的SQL查询语句。根据焊接资料库中的相关数据，分为五大类：第一类是母材，包括母材化学成分、母材组织性能、母材力学性能21个基表；第二类是焊材，包括焊条、焊丝、焊剂、保护气体24个基表；第三类是焊接方法；第四类是焊接设备；第五类是焊接坡口，第六类是焊接工艺，第七类是焊接热处理，结构如图3-5所示。

本课题以Database Desktop作为后台数据库，以Delphi7.0作为前端开发工具，基于数据挖掘原理进行焊接CAPP工艺设计系统的研究开发。

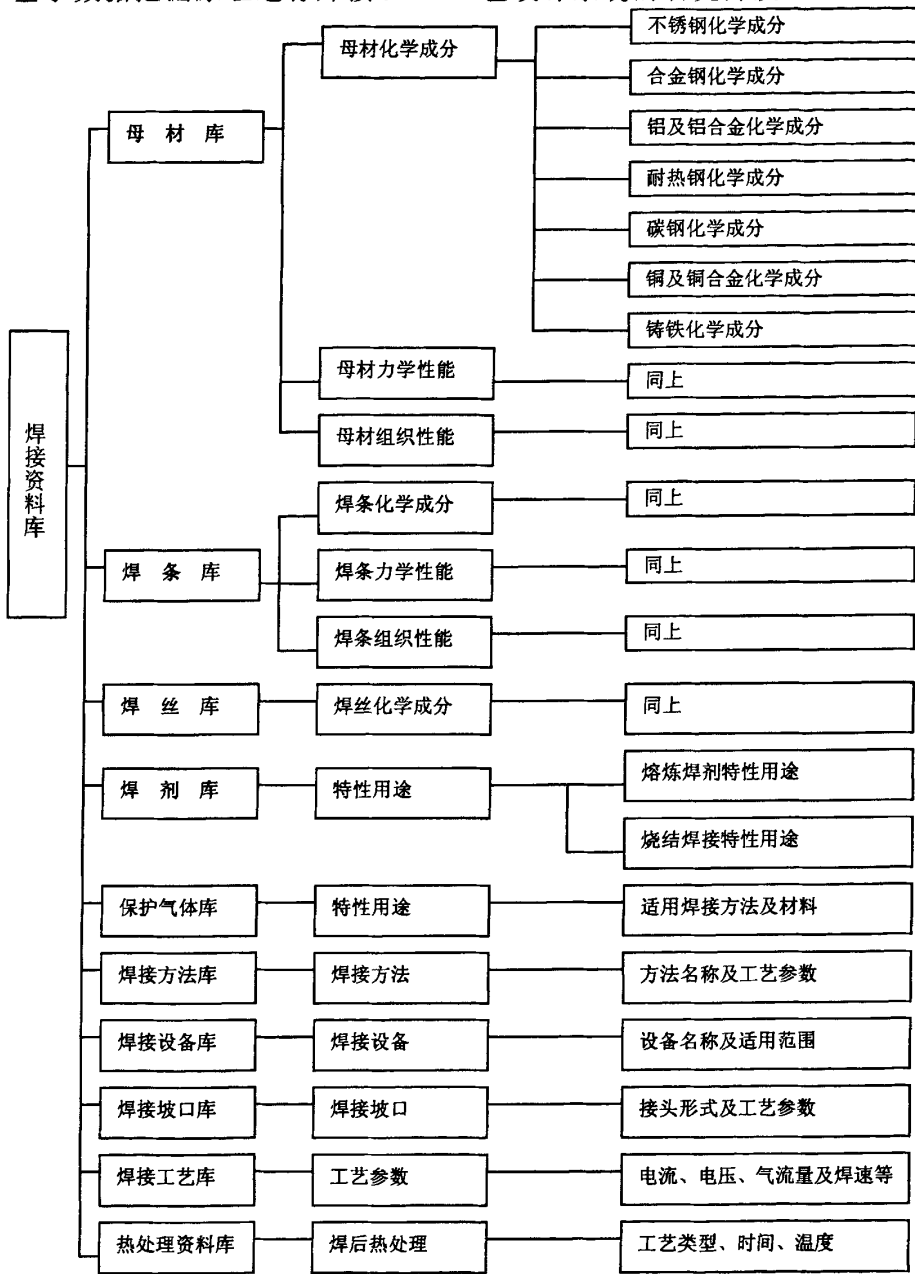


图3-5 系统数据库结构图

3.4 系统开发工具的选择

3.4.1 以 WindowsXP 作为操作平台

Windows 系统是一个用于个人计算机的交互平台。作为流行的微机图形操作系统，Windows 具有漂亮、直观、统一的用户界面和便捷的操作方式，可以

更好地发挥屏幕直观表达信息的特点。Windows 提供的操作对象只有窗口、菜单、对话框和图标等，每一个对象都代表着特定的功能，并给用户以应有的提示，使得用户学习和操作起来更为简单方便。Windows 系统同时配备了许多软件开发工具。因此，本系统采用 Windows 集成环境下的各种软件工具作为 CAPP 系统的开发平台，不但可以进一步增强系统的实用性和扩展性，使人机界面更加友好、生动，降低系统开发的成本，还能使 CAPP 系统各子模块间的动态数据交换成为可能。

3.4.2 以 Delphi 作为开发语言

Delphi 是 Borland 公司的通用软件编程工具，它的前身是 Object Pascal，自 Delphi1.0 以来，相继出现了 Delphi 的 2.0、3.0、4.0 等版本，各版本均向后兼容。随着版本的逐步提高，Delphi 在许多方面都作了改进。总之，Delphi 以其友好的可视界面，简洁明快的语言，极其丰富的第三方控件和强大的数据库及网络支持，迅速风靡世界。

在众多前台开发工具中，本系统选择了 Delphi，主要是因为 Delphi 提供了基于 Windows XP 应用程序最迅速、最简捷的方法，并具有多种数据库功能和 Internet 应用开发工具，使得开发者能够在这种交互式环境中轻松地开发各种应用程序。Delphi 具有的显著特点恰好是迅速开发 CAPP 系统所需要的条件。

(1) Object Pascal 是功能强大的面向对象语言，它的一个重要特征就是具有灵活的继承性。利用这个特性可以迅速建立加工工艺原型，然后利用 Delphi 强大的编辑功能，集成调试和测试能力，可以方便地使开发人员和用户在原型的基础上反复调整、修改和扩充原型；

(2) Delphi 开发环境可视化，开发人员可以方便地使用鼠标操作、设计、交互检测和调试 client/Server 应用程序，简化编程。利用 Delphi 开发环境，可以迅速建立一个图文并茂、色彩丰富、操作方便的人-机交互窗口图形 CAPP 系统。

(3) 为了方便用户进行数据库应用程序的开发，Delphi 提供了一个集成开发环境。包括 BDE Administrator, SQL Explorer, Database Desktop 等工具。利用这些工具，用户可以方便地设置数据库驱动程序参数、简表以及 SQL 查询。其中利用 Database Desktop，可以构造新的数据库表，修改现存数据库的字段，浏览数据库应用程序，为数据资源生成别名，并可通过输入查询表达式来查找数据库信息，而不必通过其他数据库软件另外设计数据库。

Delphi 数据库的工作过程可以简单描述为：数据库应用程序首先利用 Delphi 提供的数据库访问部件与 Borland 数据引擎 BDE 建立联系，然后通过 BDE 与数据库工具和部件、Delphi 数据库应用程序与 BDE、数据源之间建立联系^[36]。

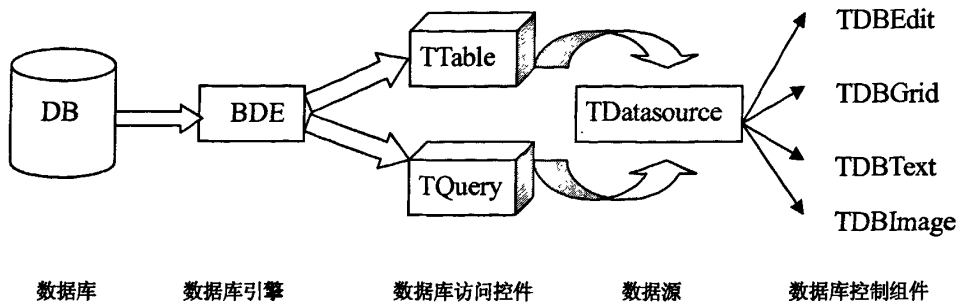


图3-6数据库组件体系结构

3.4.3 硬件开发环境

CPU: Pentium133以上

内存: 256MB以上RAM

硬盘: 40GB以上容量

显示驱动: 256色、610*480分辨率

支持Windows的显示适配卡

DVD-ROM驱动器

鼠标及其它设备

3.5 本章小结

本章详细介绍了系统作为创成式焊接 CAPP 系统的设计原理，根据汽车焊接工艺设计的特点，结合企业的生产需要，有针对性的提出了系统的整体设计思想、系统所具备的功能以及系统的结构设计内容。为了实现系统的设计功能，本章还详细介绍了焊接工艺数据库的设计流程，包括概念结构设计、逻辑设计以及物理设计，建立了为满足系统需要搭建的焊接工艺数据库。最后，系统对开发工具的选择做了简单说明，包括操作平台的选择，程序开发语言的选择以及硬件开发环境的要求。

第四章 系统模块设计与运行实例

结构的焊接过程是结构建造的关键过程，在此过程中存在着大量的焊接质量记录，这些记录是焊接结构制造过程的完整质量证据，并能清楚地体现焊接结构满足相关标准规范要求的作用。首先，结构的建造质量在相当大的程度上取决于焊接质量，而影响焊接质量的因素非常多，如钢材及焊接材料的成分、性能、焊接工艺过程、焊缝的探伤过程及探伤结果等；其次，对焊接质量记录按规定收集、储存、传递、处理和分析是质量保证体系的一个重要组成部分^[37]。

随着计算机网络化技术的发展，数据库管理系统在焊接领域，特别是在焊接生产过程中已经得到了广泛的应用。基于数据挖掘技术的创成式焊接 CAPP 系统是为了方便设计人员快捷地查询焊接工艺的相关参数，缩短工艺设计周期，快速制定工艺方案而研究开发的。系统在常规程序实现部分，通过对焊接资料库的分类，对工艺设计所需工艺参数的归类、整理，使用户可较快地查询到设计所需要的数据，并能够轻松引用大量的焊接知识和相关数据；在人工智能部分，通过编程完成数据挖掘的决策推理过程，实现对海量工艺数据潜在工艺规则的提取和挖掘过程^[38]。基于软件的功能和要求，系统以数据库系统中的表、查询作为后台存储载体及数据源，以菜单调用窗体作为前台现实载体，实现对焊接工艺数据库的查询、维护以及数据挖掘功能。程序设计作为系统最主要的部分，应当根据系统的需求分析和用户使用的具体要求来划分功能模块。根据各功能模块的要求，具体制定程序设计的流程框图，并按照框图编制的相应程序，最终完成系统设计任务，实现用户所要求的功能。

创成式 CAPP 系统是基于 WindowsXP 操作环境下的应用程序。硬件系统采用 PC 机系统模式，以单机用户为使用对象。在结构设计上，采用面向对象的模块化设计准则来达到系统所要求的设计目标。软件开发流程如下图 4-1 所示。

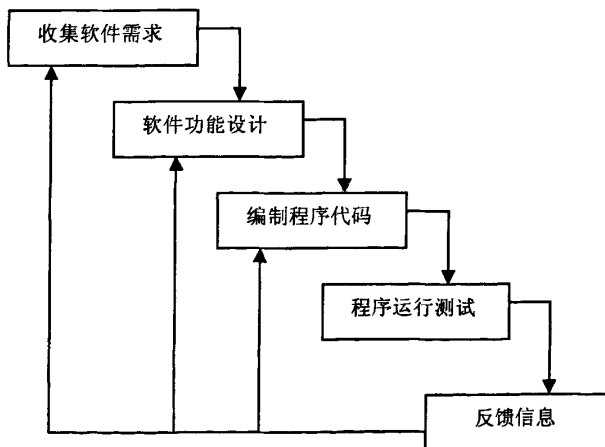


图4-1系统软件开发流程

4.1 用户登陆模块

用户打开焊接工艺数据库管理系统后,首先出现在用户面前的是系统封面,如图 4-2 所示和登录窗口。点击“进入”按钮后弹出登陆窗口,在此窗口中要求用户输入用户名和密码,然后点击“确定”按钮,系统将进行用户名和密码的确认。在登陆界面设置用户名称和密码两个 Text 控件,当用户输入的值与设置一致时,登陆系统成功,进入各功能模块选择界面,点击“退出”按钮便可退出系统。当用户输入的值与设置不一致时,验证失败,退出系统。系统的用户名和密码是以系统文件的形式保存在系统文件里的。每次启动时,从系统文件中调用用户名和密码与用户输入的密码进行比较,通过用户名称和密码验证可以防止无使用权限的用户非法进入系统模块,从而保证了数据库的安全性和完整性。界面如图 4-3 所示。

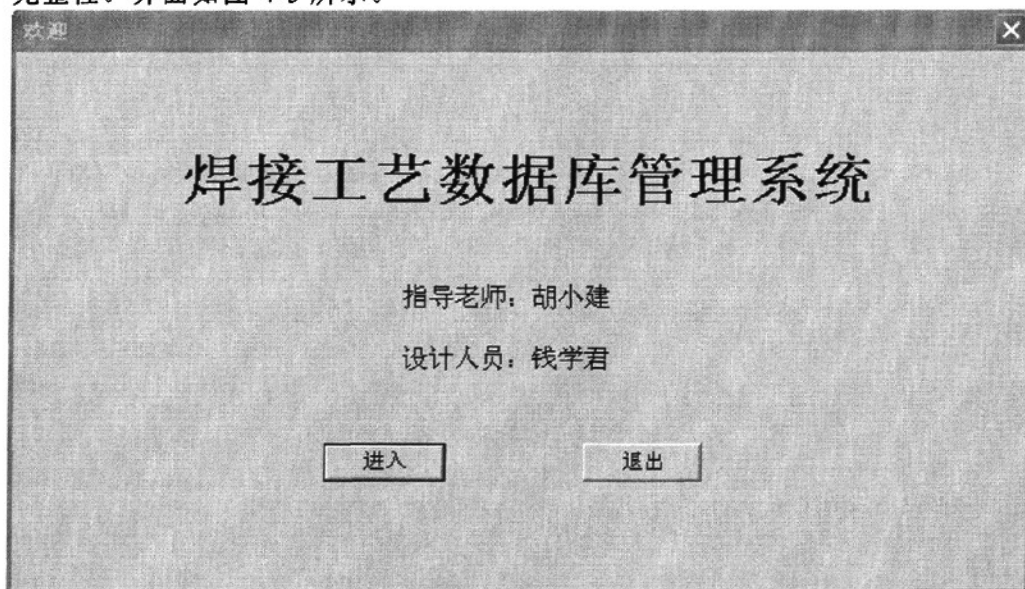


图4-2系统封面

系统通过如下的程序代码来判断用户是否为合法用户:

```
procedure TForm3.Button1Click (Sender: TObject) ;
var strID, strPWD: String;
begin
strID := edit1.Text;
strPWD := edit2.Text;
if (strID = 'qxj') and (strPWD = 'qxj') then
begin
Form3.Hide;
Form2.ShowModal;           // 显示查询界面
Form3.Close;
```

```
end
else
begin
MessageBox (handle, '用户名或密码错误, 请重新输入!', '焊接数据库及管理系统', MB_OK);
Edit1.Text := '';
Edit2.Text := '';
end;
end;
procedure TForm3.Button2Click (Sender: TObject);
begin
Form3.Close;           // 关闭登陆对话框
end;
```

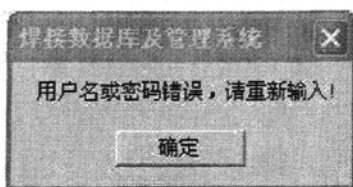
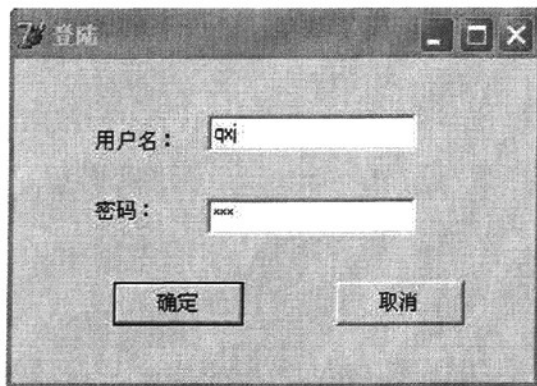


图4-3登陆窗口

4.2 系统用户界面设计

用户界面是用户与计算机之间传递、交换信息的媒介, 是用户使用计算机系统的综合操作环境。通过用户界面, 用户向计算机系统提供命令、数据等输入信息, 这些信息经计算机系统处理后, 又通过用户界面把产生的输出信息送回给用户。用户界面的核心内容包括显示风格和用户操作方式, 它集中体现了计算机系统的输入输出功能, 以及用户对系统的各个部件进行操作的控制功能。

焊接工艺数据库管理系统用户界面的设计以最大限度地满足用户的需求和使用方便为目的，在设计过程中遵循以下基本原则：

(1) 系统设计具有 Windows 环境下界面风格的用户界面，以简单明了的方式与用户进行交流，在窗口中尽可能地包含较多的信息。

(2) 用户和系统交互时，尽可能地方便用户操作，对用户的输入具有一定的容错性和检测能力，并用信息窗口来提示用户操作的注意事项和操作的正确性。

(3) 在满足功能的前提下，尽量采用按钮操作来减少数据的输入量，加快数据的录入速度，提高系统解决问题的能力。

(4) 在同一用户界面中，用窗口分隔不同种类的信息，所有的菜单选择、数据输入与显示和其它功能保持风格的一致性。

系统登陆后将出现焊接工艺数据库管理系统的全界面。文件、编辑、数据库维护、帮助等几个菜单，在所有的应用程序中都可以看到。文件菜单用于用户对系统文件的操作，提供对文件的打开、新建、退出等操作。编辑菜单提供了常用的撤销、剪切、复制、粘贴、删除、全选功能。数据库维护菜单提供首记录、上一条记录、下一条记录、末记录、添加记录、编辑记录、删除记录、确认修改、取消修改、刷新记录功能。在需要帮助时，用户通过帮助菜单可以查阅本系统帮助主题和系统介绍，如图 4-4 所示。

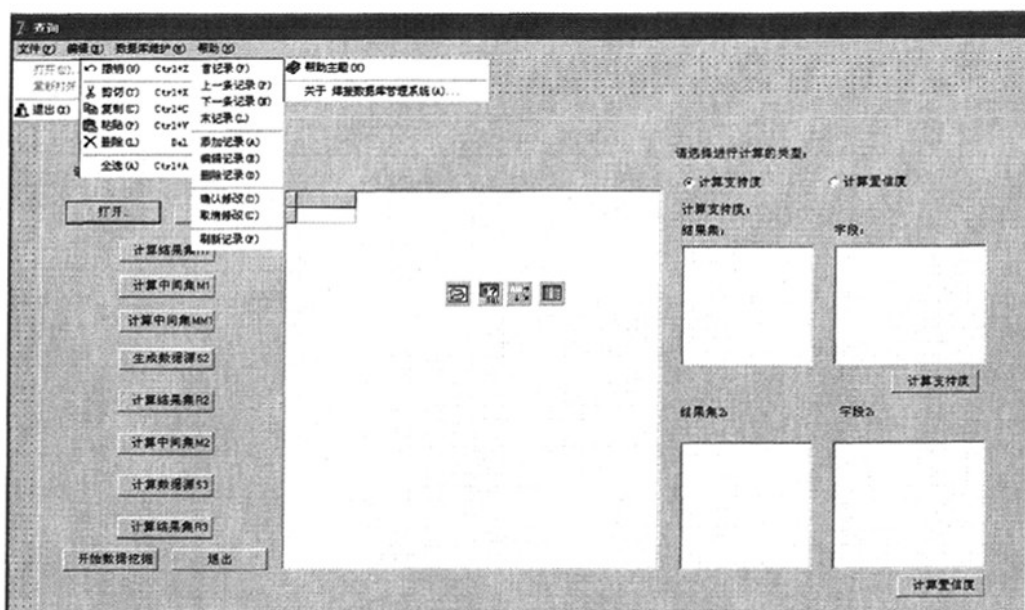


图 4-4 系统用户界面

根据生产中最常用的焊接过程，把所设计的焊接 CAPP 系统划分为 3 个主要模块，分别是系统查询模块、焊接工艺设计模块、数据挖掘模块。在系统设计时选用选项卡可以方便地切换不同模块的查询界面。系统主功能模块分为查

询模块、焊接工艺设计模块、数据挖掘模块，分别用于查询焊接资料库、坡口工艺设计和工艺信息的数据挖掘。

4.3 系统查询模块设计

此模块提供了对母材、焊材相关信息的查询。利用用户输入的条件，对焊接资料库进行查询，查询结果以表格和图片形式显示，这样可以使用户能够直观而方便地得到想要的查询结果，供用户在制定焊接工艺的过程中参考。

4.3.1 母材查询

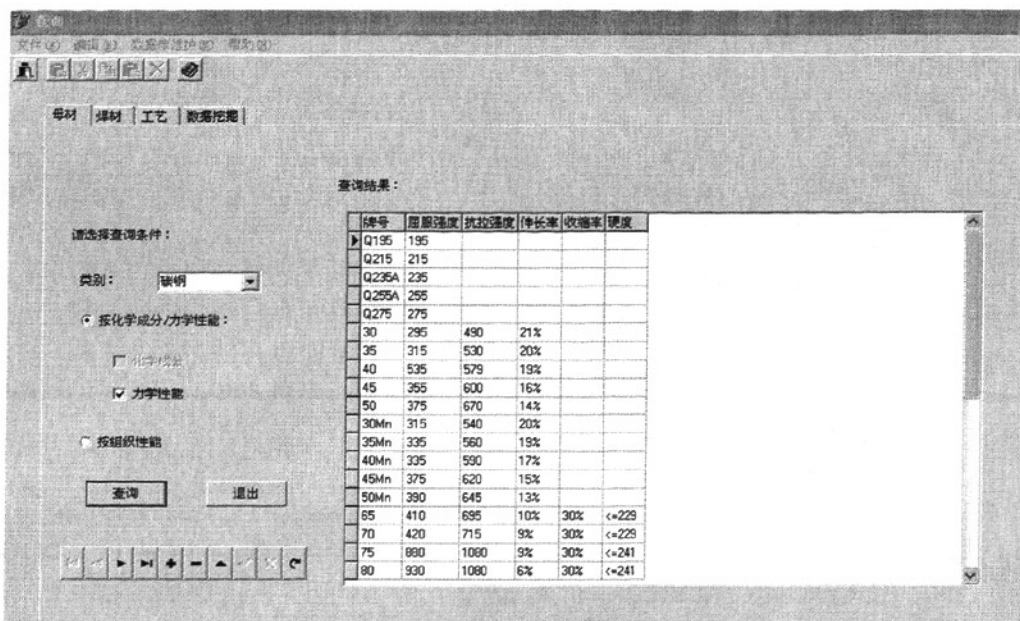
点击主界面下的“母材”选项卡，可进行母材相关信息的查询。在此界面中，母材类别中采用了ComboBox控件，用户无需手工输入母材类别，只要点击下拉列表框即可显示出所有的母材类别，包括碳钢、不锈钢、耐热钢、合金钢等。母材数据库中包含了钢材牌号、化学成分、组织性能，力学性能等方面的数据，完全能够满足工厂的生产需要。这样设计一方面减少了用户的手工输入量，提高输入效率；另一方面可以减少手工输入错误的发生，提高输入质量。

例如，首先在类别栏的下拉菜单中选择需要查询的母材种类“碳钢”，然后选中“按化学成分/力学性能：”按钮，选中“化学成分”或“力学性能”，点击“查询”即可得到想要查询的母材化学成分或力学性能（如图4-5，4-6所示）；在类别栏的下拉菜单中选择母材种类“铝及铝合金”，选中“组织性能”按钮，点击“查询”即可得到铝及铝合金的组织性能（如图4-7所示）。

查询结果：

| 牌号 | C | Mn | Si | S | P | Ni | Cr | Cu |
|-------|-----------|-----------|-----------|---------|---------|--------|--------|--------|
| Q195 | 0.06-0.12 | 0.25-0.50 | <=0.30 | <=0.050 | <=0.045 | | | |
| Q215 | 0.09-0.15 | 0.25-0.55 | <=0.30 | <=0.050 | <=0.045 | | | |
| Q235A | 0.14-0.22 | 0.30-0.65 | <=0.30 | <=0.045 | <=0.045 | | | |
| Q255A | 0.18-0.28 | 0.40-0.70 | <=0.30 | <=0.050 | <=0.045 | | | |
| Q275 | 0.28-0.38 | 0.50-0.80 | <=0.35 | <=0.050 | <=0.045 | | | |
| 30 | 0.27-0.35 | 0.50-0.80 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 35 | 0.32-0.40 | 0.50-0.80 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 40 | 0.37-0.45 | 0.50-0.80 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 45 | 0.42-0.50 | 0.50-0.80 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 50 | 0.47-0.55 | 0.50-0.80 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 30Mn | 0.27-0.35 | 0.70-1.00 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 35Mn | 0.32-0.40 | 0.70-1.00 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 40Mn | 0.37-0.45 | 0.70-1.00 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 45Mn | 0.42-0.50 | 0.70-1.00 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 50Mn | 0.48-0.56 | 0.70-1.00 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 65 | 0.62-0.70 | 0.50-0.80 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 70 | 0.67-0.75 | 0.50-0.80 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |
| 75 | 0.72-0.80 | 0.50-0.80 | 0.17-0.37 | <=0.035 | <=0.035 | <=0.25 | <=0.25 | <=0.25 |

图4-5 母材化学成分查询结果界面



4-6 母材力学性能查询结果界面

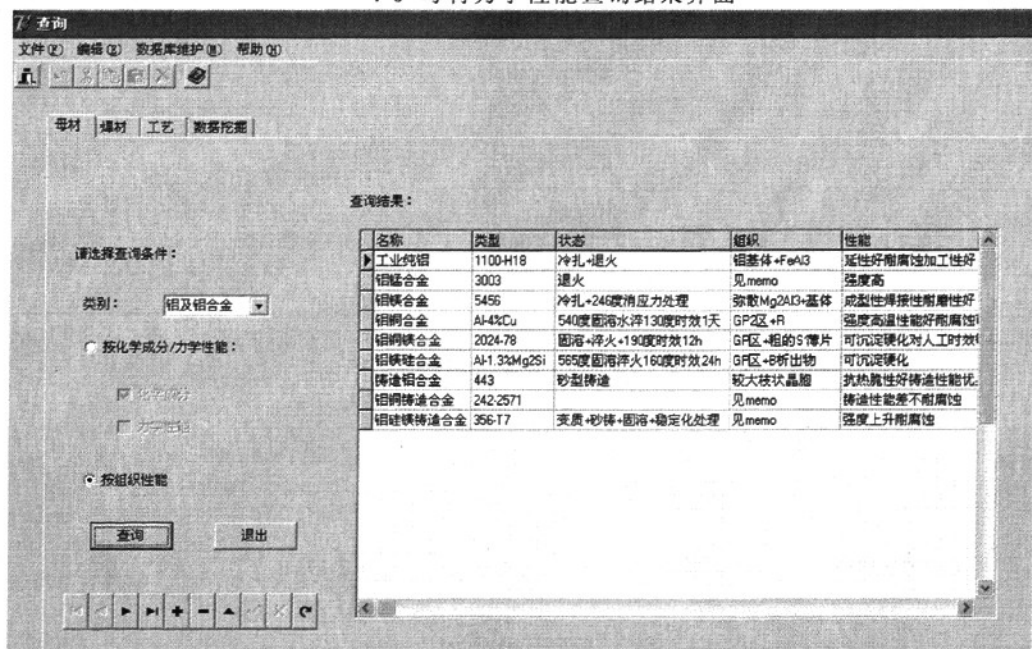


图4-7 母材组织性能查询结果界面

在界面内有十个命令按钮可供使用，它们分别是“首记录”、“上一页”、“下一页”、“末记录”、“增加一条记录”、“删除一条记录”、“编辑”、“保存”、“删除”、“更新”按钮。通过增加一条记录，可以向焊接工艺数据库中添加有用的数据记录丰富数据库。通过删除功能，可以将数据库中不需要的数据记录删除，减小数据库的容量。通过编辑功能，可以将数据库中输入不正确的数据记录进行更正。在此界面中可以对原始数据库进行浏览和维护工作。由于一个实用性

的系统开发是一个长期的使用、修改和扩充的过程，在一定时期内，焊接资料库中的信息不一定能够完全满足实际生产的需要，因此，提供给工艺设计人员修改焊接资料库的接口是非常必要的。设计人员首先搜索需要编辑的记录，点击“编辑”按钮，即可对参数进行添加、修改等操作，操作完成后，点击“保存”按钮即可完成此次操作，点击“取消”按钮取消此次工作。

4.3.2 焊材查询

点击主界面下的“焊材”选项卡，就可进行焊材相关信息的查询。主要涉及到不同类别焊条的化学成分、力学性能及特性用途的查询，不同类别的焊丝化学成分查询，保护其他的特性用途，烧结焊剂以及熔炼焊剂的特性用途查询。为防止用户的误操作，系统设置了使能功能，可以正确引导用户查询资料库中已存在的数据内容。与母材查询界面相似，焊材查询界面内也有十个命令按钮可供使用，在此对其使用方法不再赘述。

例如，首先在种类栏的下拉列表中选择“焊条”，在类别栏的下拉列表中选择“耐热钢”，选中“力学性能”按钮，点击查询，即可得到耐热钢焊条的力学性能，如图4-8所示。

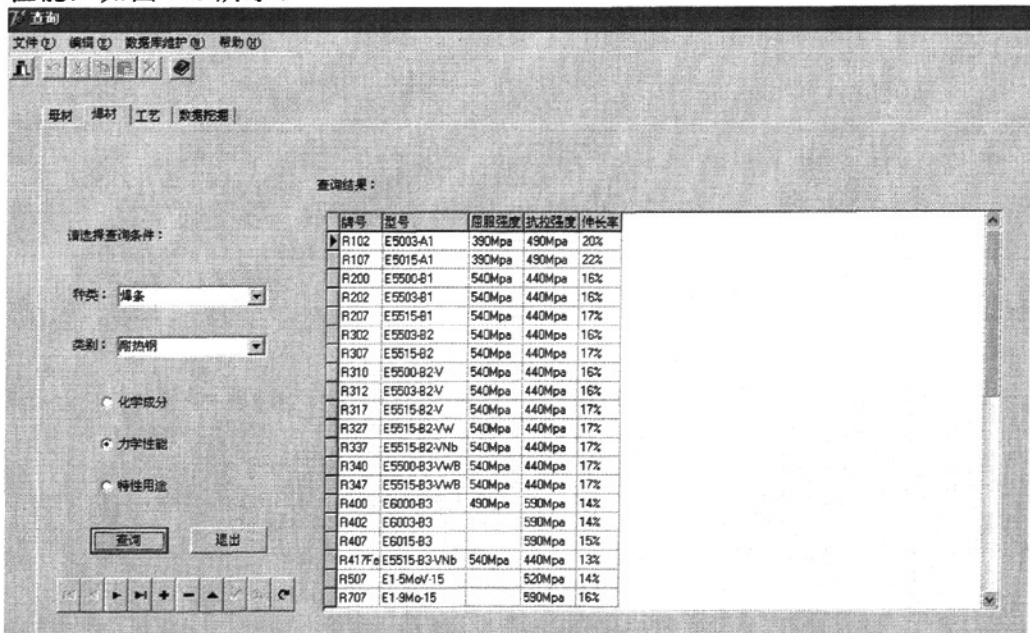


图4-8 焊条力学性能查询

在种类栏的下拉列表框中选择“焊丝”，在类别栏的下拉列表框中选择“铸铁”，此时只有“化学成分”按钮可选，即对不同类型的焊丝只可对其化学成分进行查询，如图4-9所示。

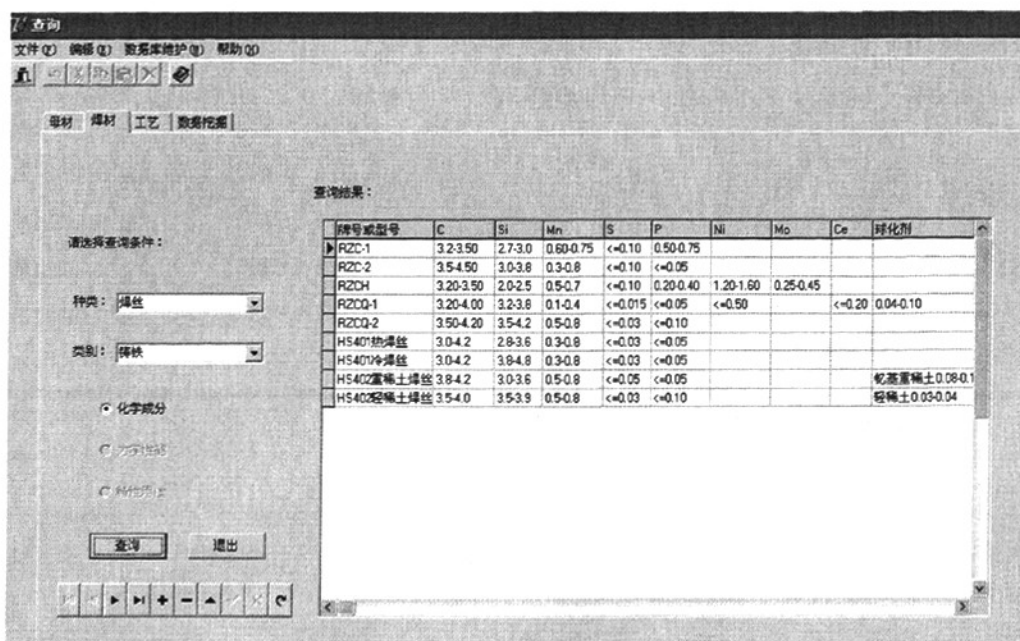


图4-9 焊丝化学成分查询

在种类栏的下拉列表中选择“保护气体”，“烧结焊剂”，“熔炼焊剂”，则无类别可选，只可对其特性用途进行查询，如图4-10，4-11，4-12所示。



图4-10 保护气体特性用途查询

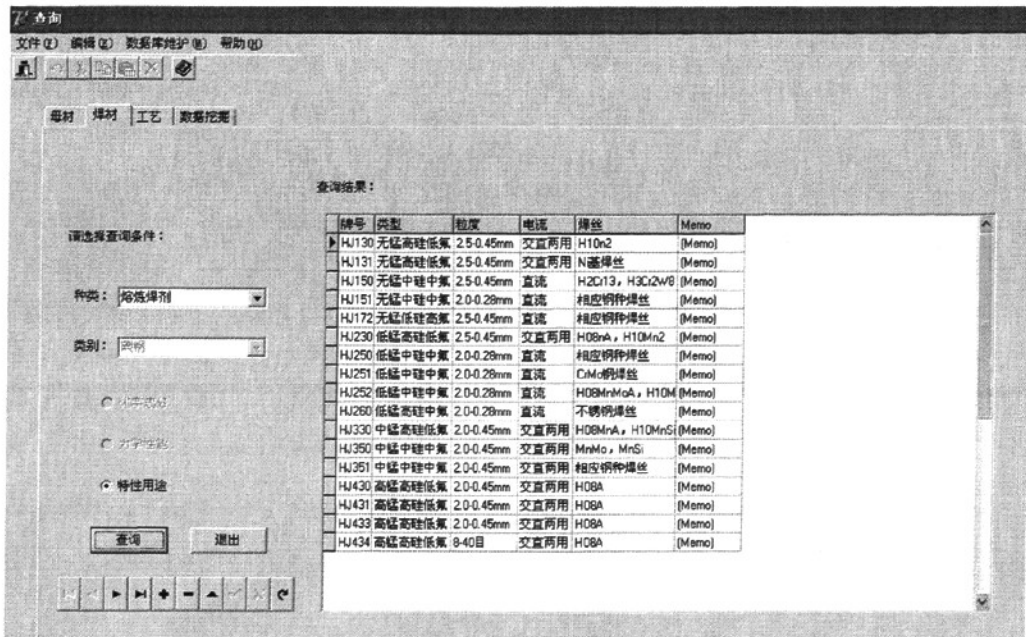


图4-11 熔炼焊剂特性用途查询

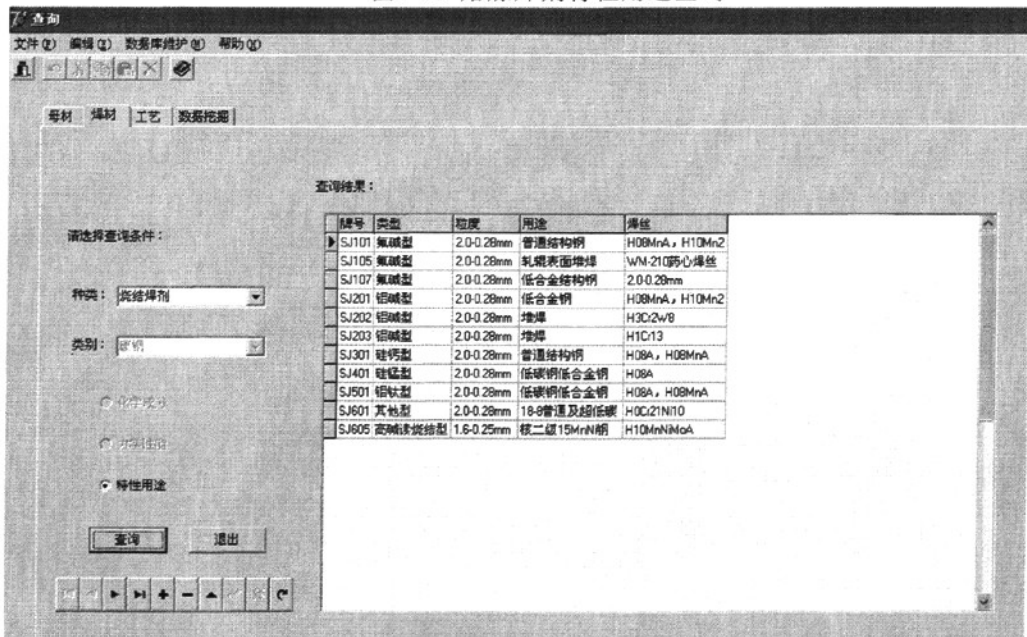


图4-12 烧结焊剂特性用途查询

由于焊接工艺数据库的维护是十分重要的操作，它涉及到焊接工艺数据库的完整性与正确性以及焊接工艺设计功能的正常实现，所以应该谨慎为之，防止因误操作而造成不必要的损失，用户需要仔细检查并确认无误后再点击“保存”按钮，最终完成焊接工艺文件的编辑。这一点对于系统的安全性非常重要。

4.4 焊接工艺制定模块

焊接坡口是保证获得优良焊接接头性能的重要措施，因此焊接坡口设计是焊接工艺设计的重要部分。不同的焊接方法和接头形式具有不同的坡口形式，在制定焊接工艺的过程中，通常根据国家标准来选择坡口形式。在本系统中设计了坡口图形库，包括坡口所属接头类型、坡口形式、坡口适用的焊接方法、母材厚度等，另外，系统还包括了坡口形状参数以及焊接电流、焊接电压和焊接速度等比较典型的工艺规范参数。由于焊接工艺规范参数的选择比较复杂，工件厚度的取值范围比较灵活，坡口形式种类繁多，以及坡口形状参数的多样化，系统在设计时采用用户输入的方式进行查询。系统在查询过程中，可以任意填入坡口的已知参数，选择单条件查询或采用组合查询的方式限定坡口工艺设计的查询范围，当参数设置完毕，点击“查询”按钮，执行查询功能。

为了将已有的工艺数据作为可靠的参考资料方便今后的坡口工艺设计，该模块除了查询功能外，还添加了数据库维护功能，即对原有的焊接坡口资料库添加新的内容。页面下方，同样设置了十个命令按钮便于使用，它们分别是“首记录”、“上一条”、“下一条”、“末记录”、“增加一条记录”、“删除一条记录”、“编辑”、“保存”、“删除”、“更新”按钮。在执行查询功能时，只可使用“下一条”和“末记录”按钮，对原始数据库进行浏览，如图4-13所示。

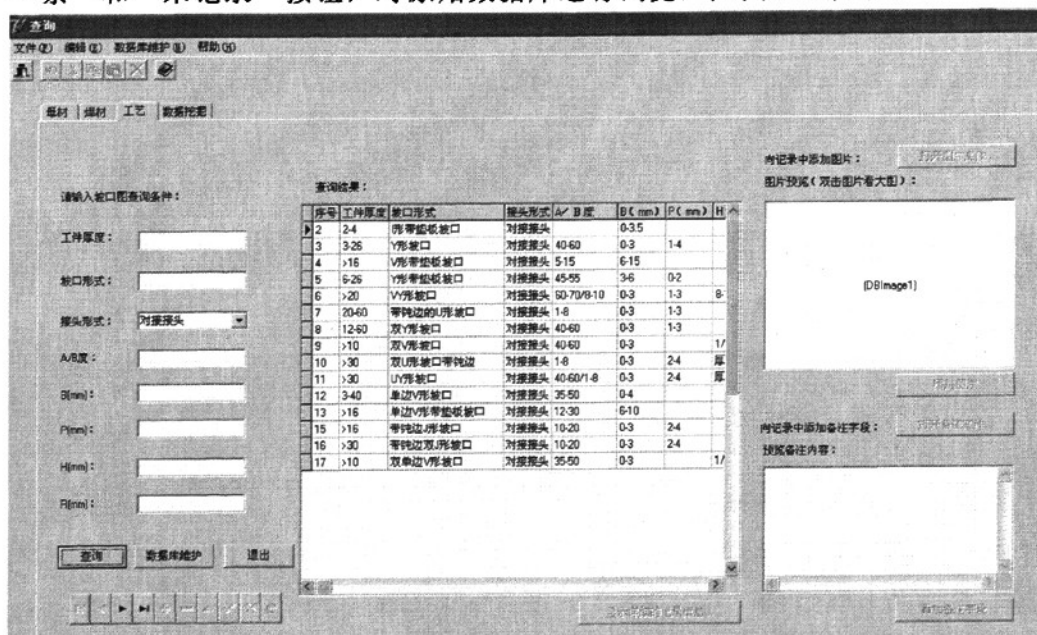


图4-13 坡口工艺查询

在执行数据库维护功能时，可使用“编辑”，“删除”等按钮进行操作，完成对工艺数据的添加、编辑、修改和删除等操作。同时，“打开图片文件”和“打开备注文件”按钮使能，用户可以向日记录中添加或更换满意的坡口示意图或备注字段。点击“打开图片文件”按钮，系统提示默认的存储文件夹，选中图片

在图形显示框中进行图片预览，双击该图片，系统会自动弹出放大的图片，便于用户浏览，如图 4-14、4-15 所示。点击“添加图片”按钮可将图片添加到指定的字段中去。同样，点击“打开备注文件”按钮，选择预备添加的工艺文件，在预览备注内容显示框中进行预览，点击“添加备注字段”按钮将预备添加的文字添加到备注字段中，如图 4-16 所示。

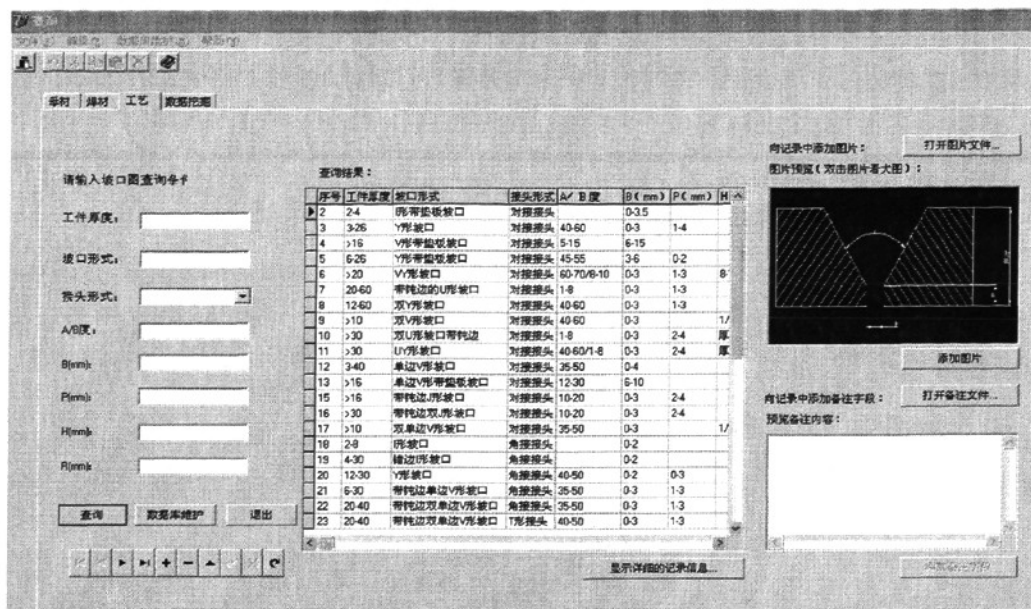


图4-14 坡口工艺设计数据库维护1

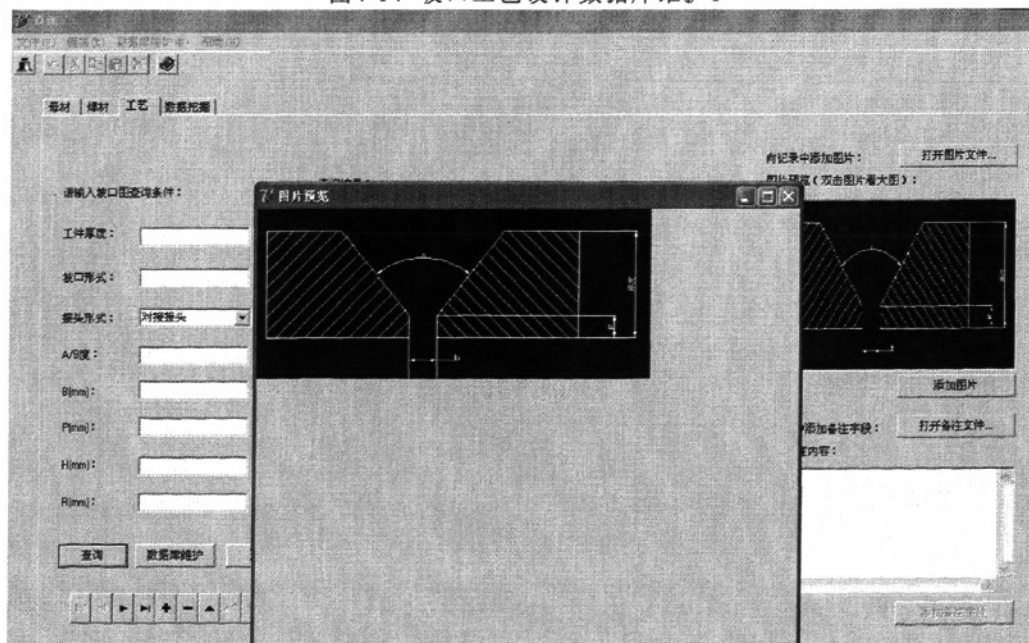


图4-15 坡口图片预览



图4-16 坡口工艺设计数据库维护2

在坡口工艺设计模块，根据查询条件可以显示出符合条件的查询结果，单击“显示详细的记录信息”按钮进入下一个页面，在这个页面上，显示了该条信息的所有工艺参数、坡口示意图以及备注，如图 4-17 所示。

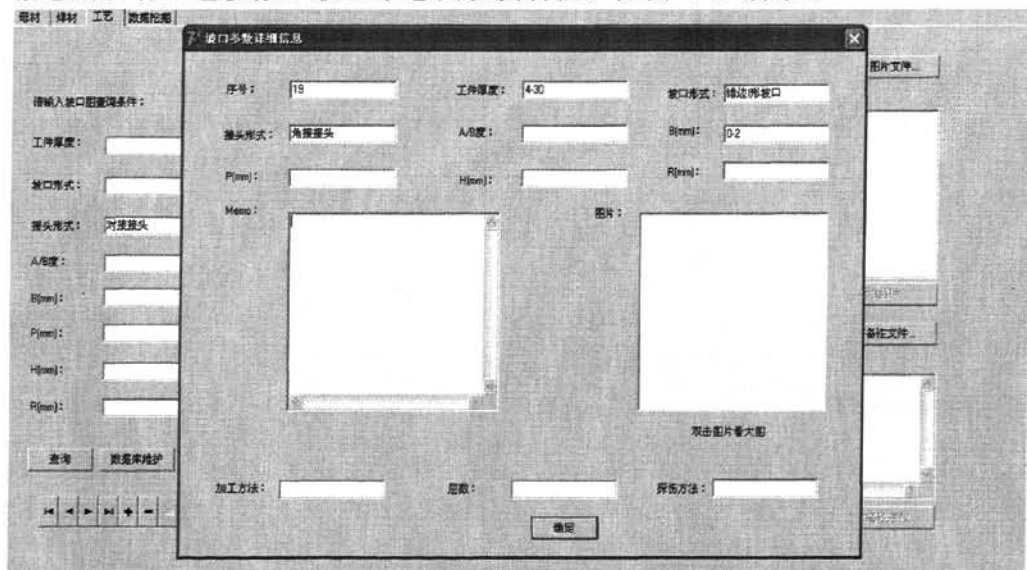


图4-17 详细信息示意图

4.5 数据挖掘模块

本系统的智能体现在运用数据挖掘技术实现对潜在工艺规则的提取。在制定焊接工艺时，数据挖掘算法将根据用户输入的工艺文件数据进行处理，在与用户的交互过程中最终实现焊接工艺规则的提取。由于关联规则算法是目前数

据挖掘技术中较为成熟的算法以及简单易行的特点,所以选择 Apriori 改进算法作为本课题所建立数据挖掘的训练算法。在系统主界面上单击“数据挖掘”选项卡,点击“打开”按钮,页面提示打开需要进行处理的焊接工艺文件,选择“测试数据 1”,出现图 4-18 所示的规则获取页面。



图4-18 测试数据选择界面

点击“计算结果集 R1”按钮,系统执行 Apriori 改进算法,将数据表中的字段转化为一维项集,并对一维项集的每一项进行计数,提取计数大于最小支持度的项,并将符合条件的项和计数保存在结果集 R1 中,如图 4-19 所示。

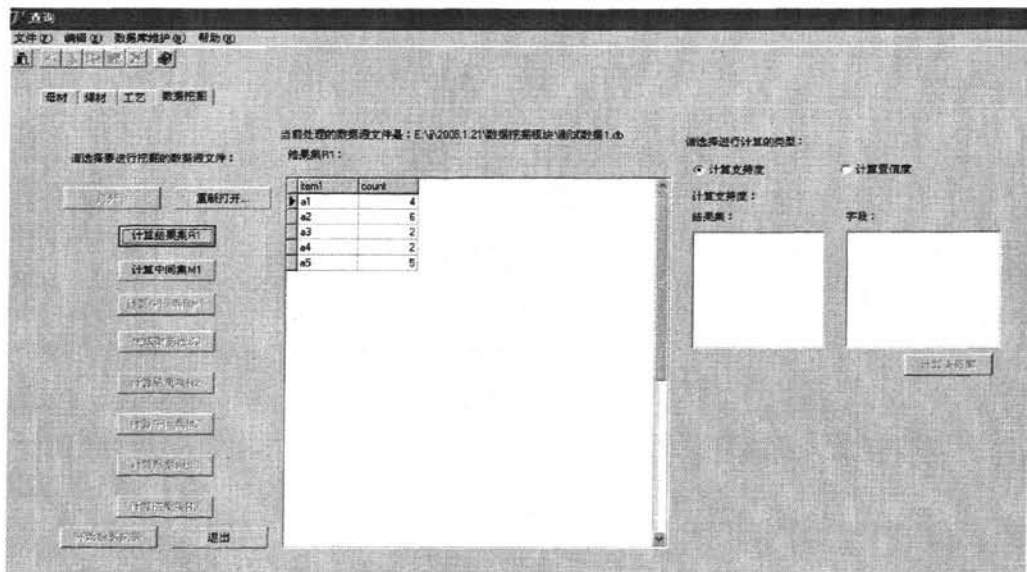


图4-19 结果集R1

点击“计算中间集 M1”按钮,系统将结果集 R1 与测试数据源连接,将数

据源中低于最小支持度计数的项删除，显示大于等于最小支持度计数的项，以及该项在数据源中的 id，将结果保存在一维中间项集 M1 中，如图 4-20 所示。

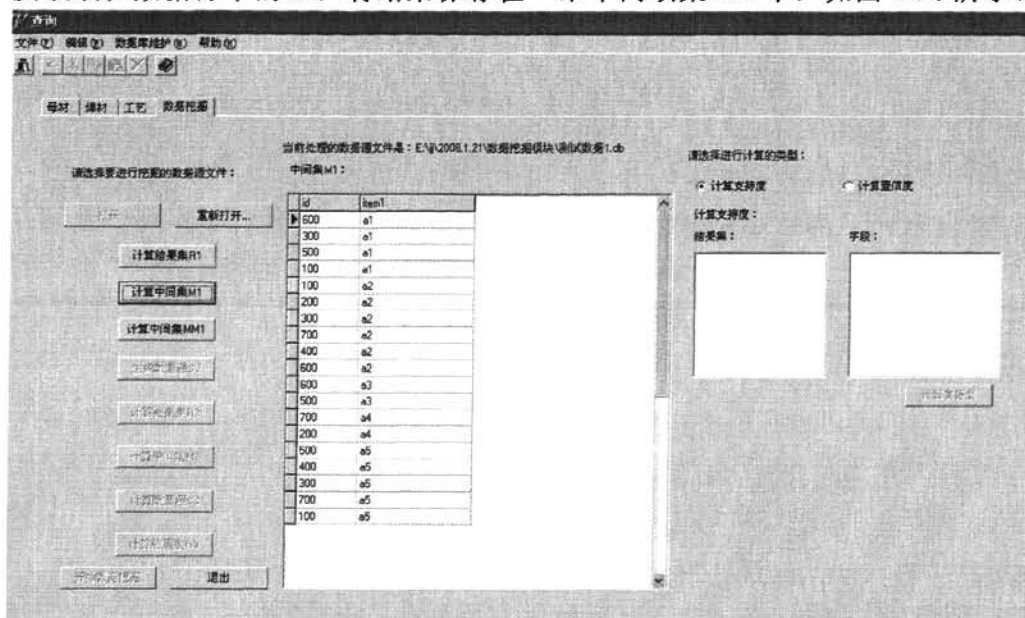


图4-20 中间项集M1

点击“计算中间集 MM1”按钮，对中间项集 M1 的 id 和 item1 进行拷贝，将复制结果保存在中间项集 MM1 中，如图 4-21。

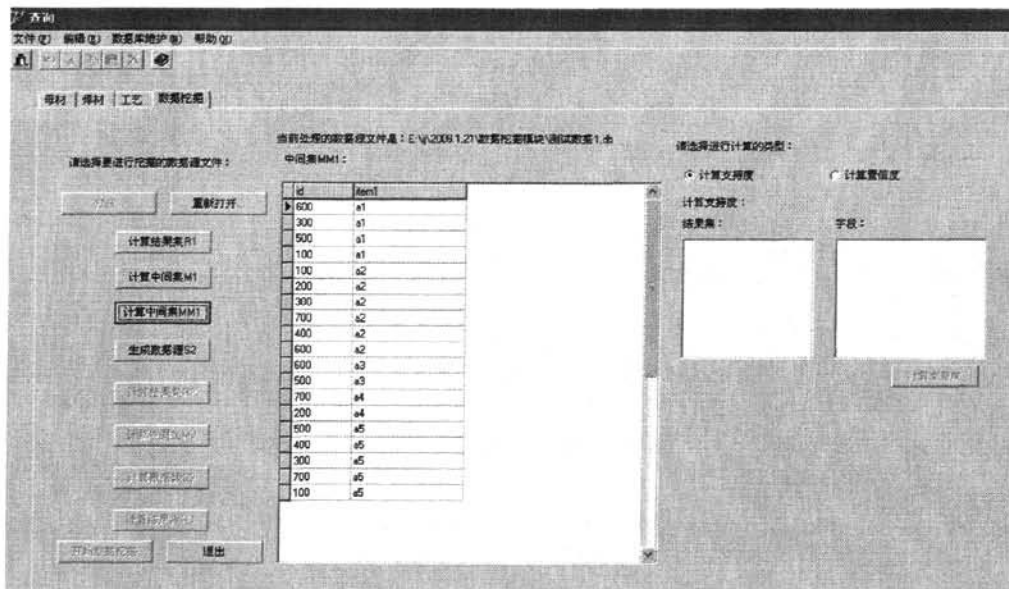


图4-21 中间项集MM1

点击“生成数据源 S2”按钮，当中间项集 M1 与项集 MM1 的 id 相同，而 item1 不相同，两个数据表进行自交，产生二维数据源 S2，如图 4-22 所示。

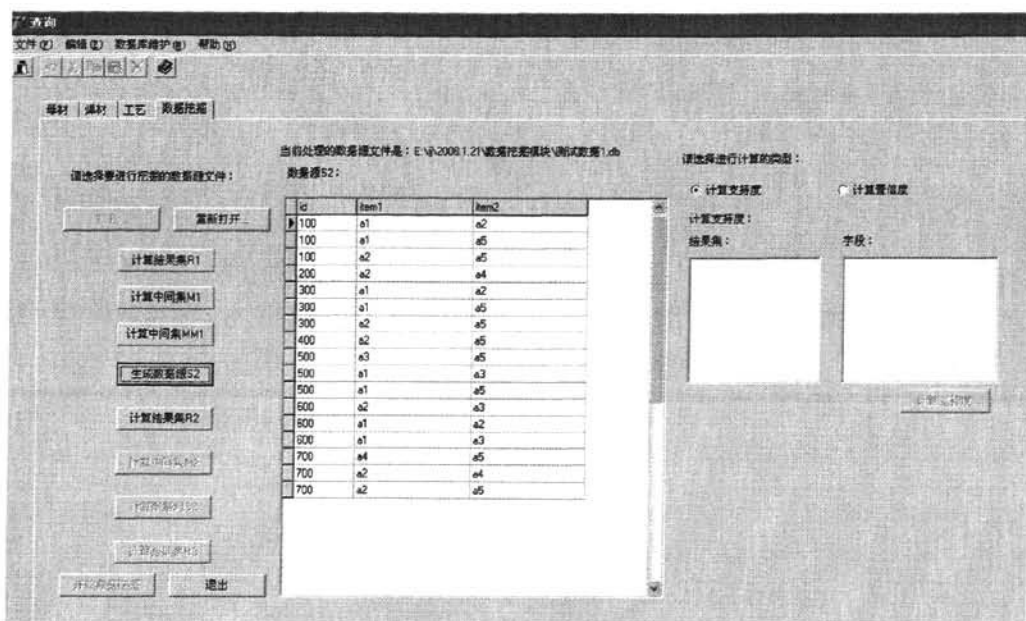


图4-22 数据源 S2

点击“计算结果集 R2”按钮，当数据源 S2 每条记录中 id 不同，而 item1 与 item2 相同时，对该记录进行计数，提取支持度大于设置的最小支持度的项集，并显示计数，保存在结果集 R2 中，如图 4-23 所示。

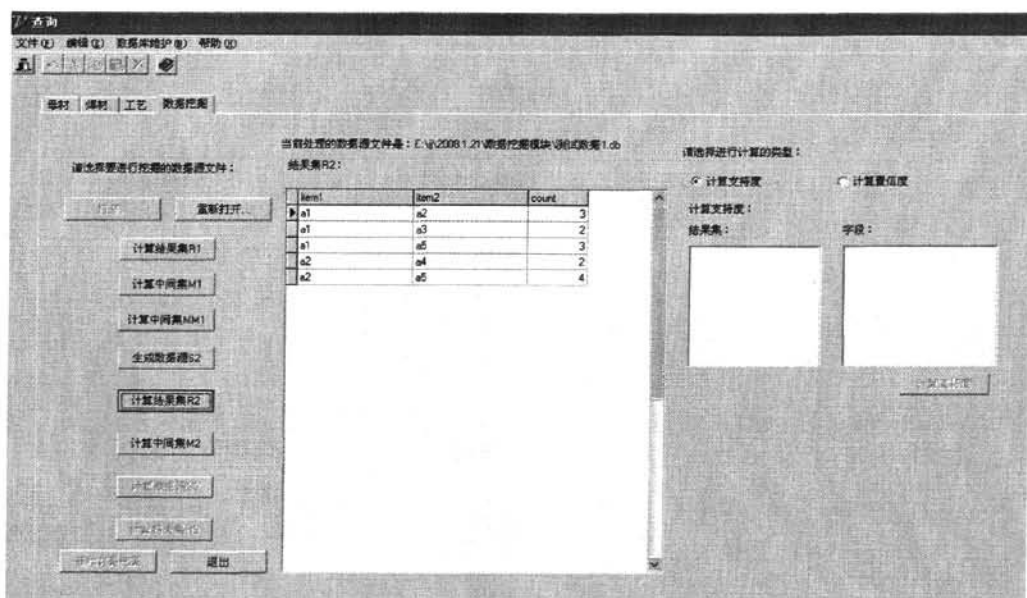


图4-23 结果集 R2

点击“计算中间集 M2”按钮，将计算的结果集 R2 与数据源 S2 进行比较，提取 R2 中出现的记录在数据源中的 id，并将 id, item1, item2 保存在中间集 M2 中，如图 4-24 所示。

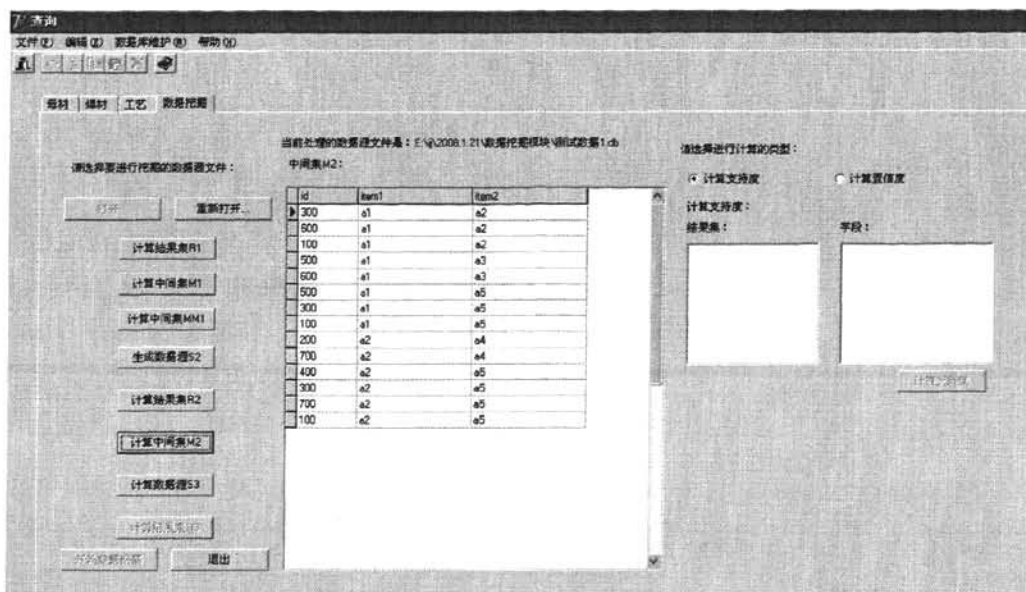


图4-24 中间集M2

点击“计算数据源 S3”按钮，当中间项集 M2 与项集 MM1 的 id 相同，而 item 不相同，两个数据表进行自交，产生三维数据源 S3，如图 4-25 所示。

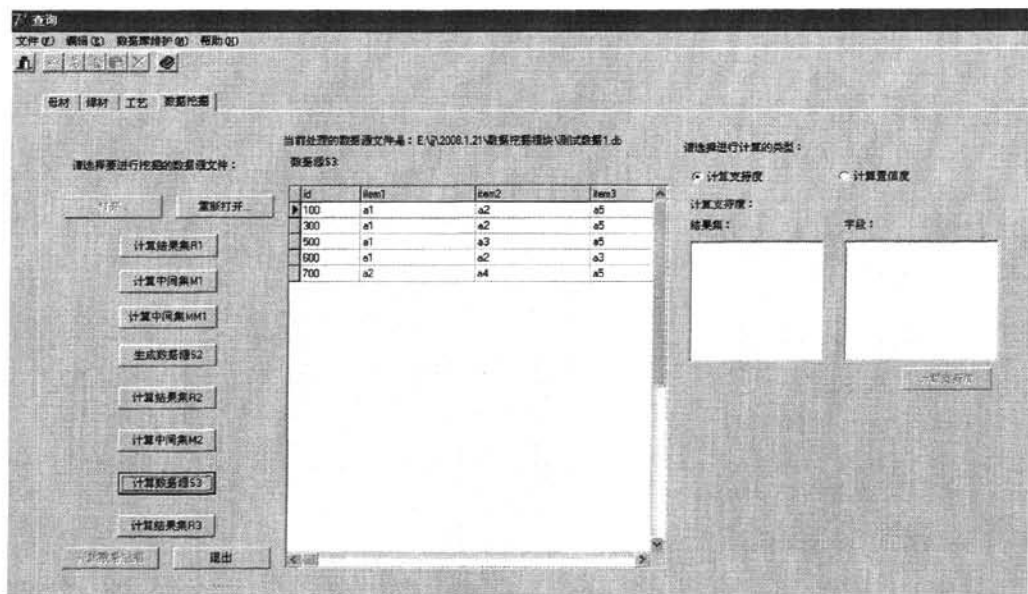


图4-25 数据源S3

点击“计算结果集 R3”按钮，对数据源 S3 不同 id 中相同的记录进行计数，当计数值大于初始设置的支持度计数时，提取记录并将计数保存在结果集 R3 中，如图 4-26 所示。

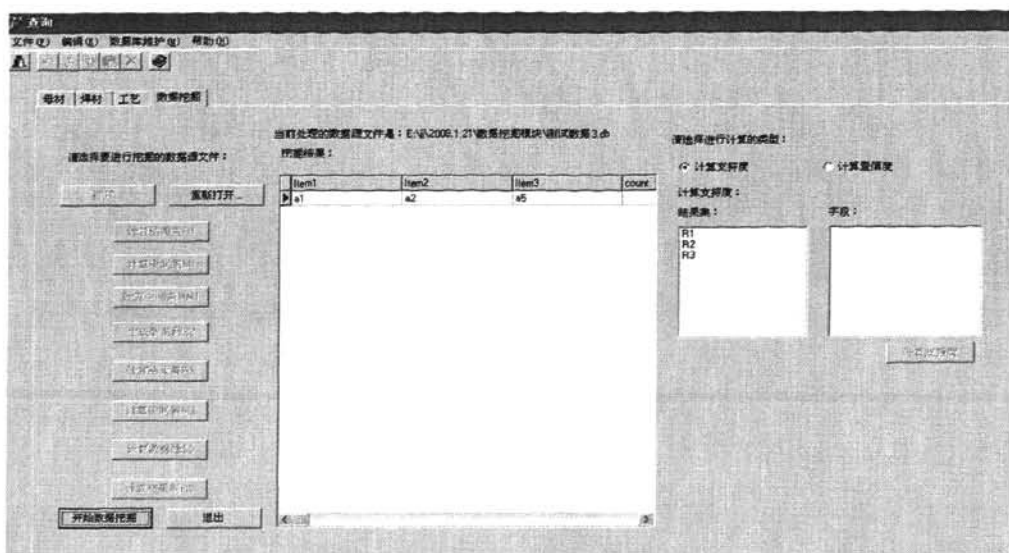


图4-26 结果集R3

当系统对测试数据挖掘到结果集 R3 时，满足支持度计数的记录只有 1 条，计数为 2，挖掘算法结束，得出满足最小支持度计数的项集为 {a1, a2, a5}。

点击“开始数据挖掘”按钮，系统默认进行支持度计算，在结果集栏内显示频繁一项集，频繁二项集和频繁三项集。鼠标任意点击“R1”，“R2”或“R3”，字段栏内会向用户提示频繁项集中的所有记录。以选择计算频繁二项集 R2 中 a1 和 a5 的支持度为例，在结果集中选中 R2，在字段栏中系统会显示出所有的频繁二项集，选中 a1 和 a5 所在的行，点击“计算支持度”按钮，显示支持度是 42.86%，此例说明，项 a1 和 a5 同时出现的概率为 42.86%，如图 4-27 所示。

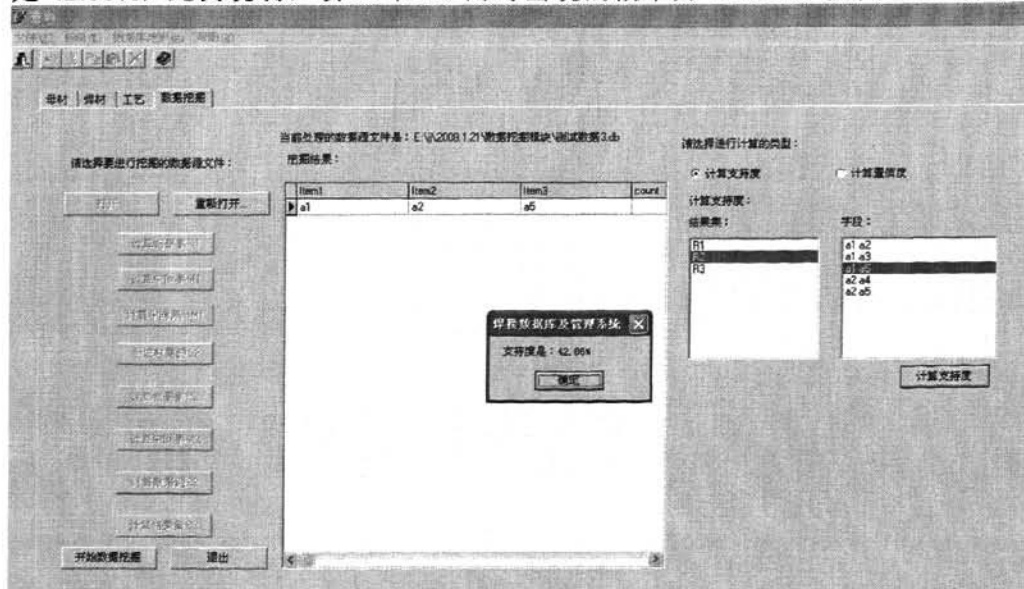
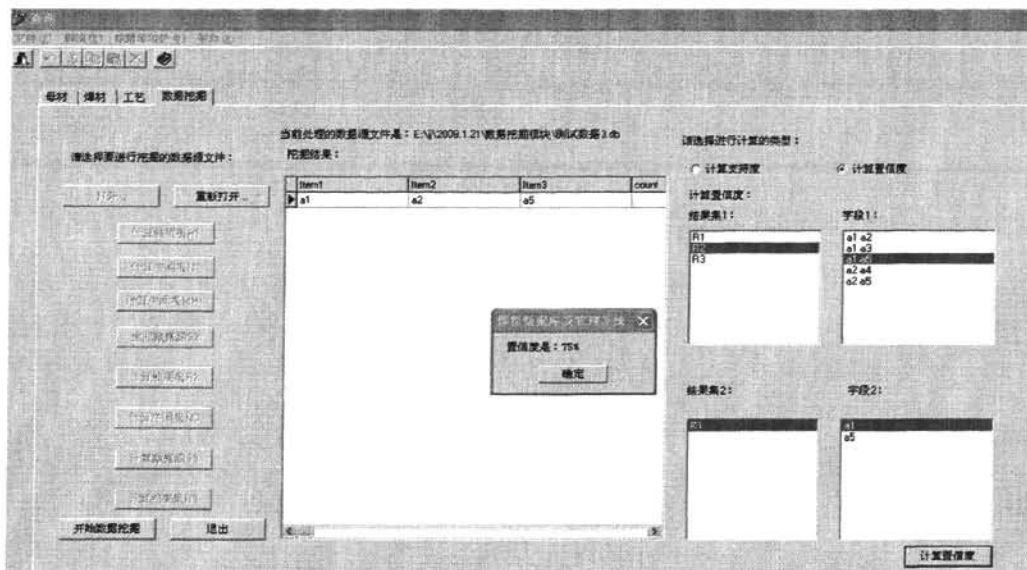


图4-27 支持度计算

支持度计算完毕，点击“计算置信度”选项，系统提示计算置信度所需的

显示框，对于频繁二项集的置信度计算，则是对其频繁一项集进行计算，在结果集 2 显示框中，提示只有 R1 可选，点选后 R1 后，在字段 2 显示框中提示有 a1, a5 两项可供选择。本例选中 a1，点击“计算置信度”按钮，系统自动弹出计算结果显示框，提示置信度是 75%。此例的演示结果说明，在 a1 出现的情况下，a5 出现的概率为 75%，a1, a5 是强相关的，如图 4-28 所示。



4-28 置信度计算

上述文字和图片展示了焊接 CAPP 系统数据挖掘的原理和操作过程，实际生产中，点击“重新打开”按钮，选择离线的工艺文件后，点击“开始数据挖掘”按钮，即可对工艺参数进行支持度和置信度的计算。如图 4-29, 4-30 所示。

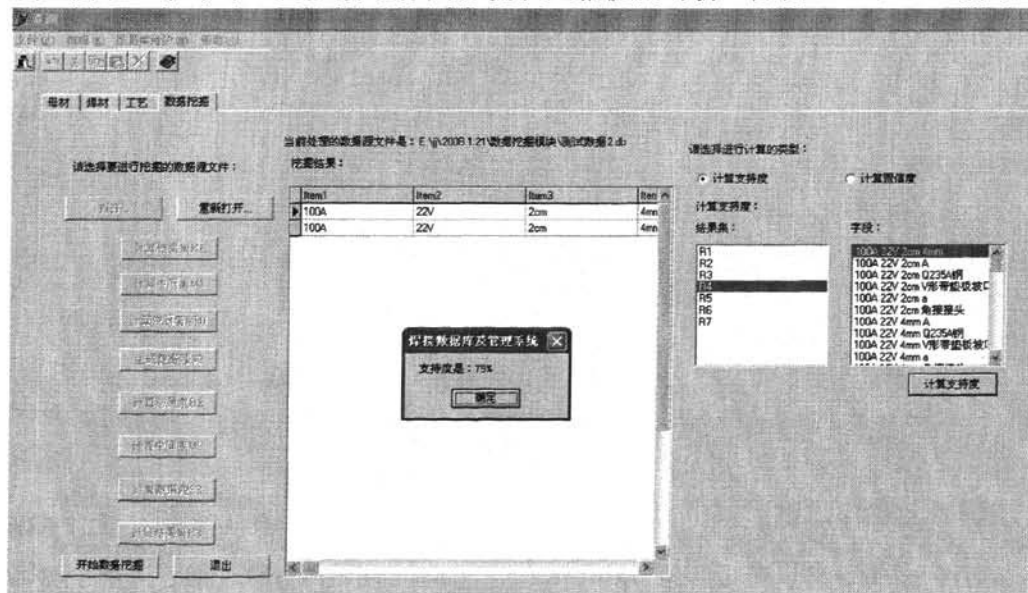


图4-29 焊接工艺参数的支持度计算

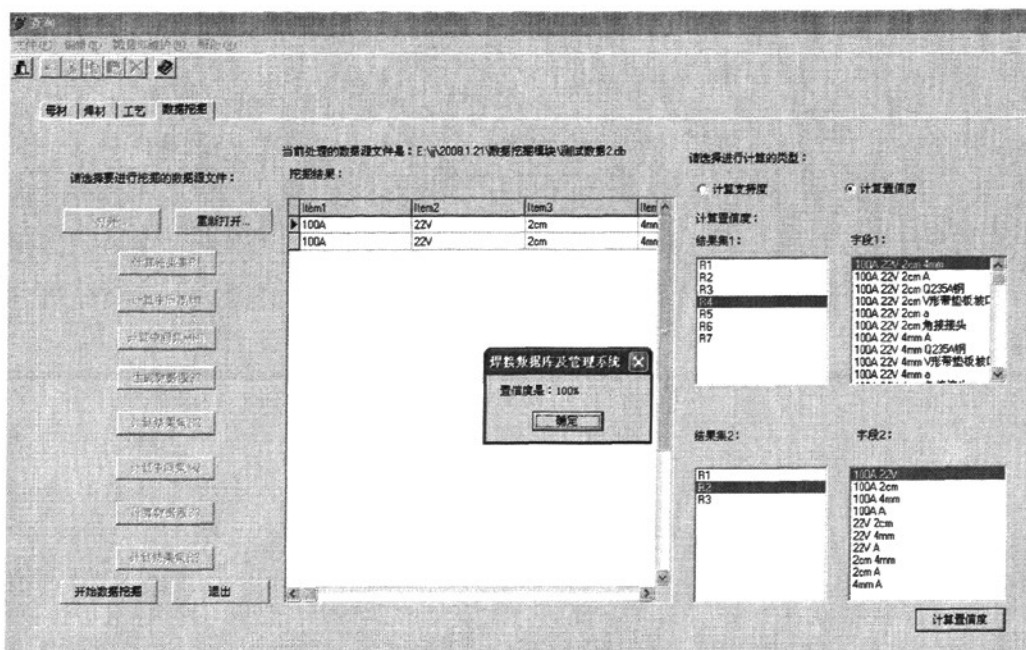


图4-30 焊接工艺参数的置信度计算

此例的挖掘结果说明：板厚为 4mm，焊接电流为 100A，焊接电压为 22V，焊丝伸长度为 2cm 的情况，在整个工艺文件中出现的概率为 75%；当电流为 100A，电压为 22V 的情况下，焊丝伸长为 2cm，板厚为 4mm 的出现概率为 100%。工艺人员在设计工艺流程或工艺参数时，根据工厂已有的大量工艺数据，在对其进行适当的整理后，可以直接利用挖掘出的潜在关联规则指导新工艺的设计，提高焊接工艺设计水平和效率，缩短工艺编制的时间，优化焊接工艺流程。

焊接过程是一个复杂的过程，很多参数难以量化，存在许多不确定的模糊知识，往往需要根据经验来做决定。因此，在焊接工艺设计过程中采用数据挖掘技术，不仅具有精度高、速度快，在许多方面比人类专家更为准确、迅速而无遗忘的特点，系统还容易复制和移植，经验易于传播。一方面可以解决工艺工程师的匮乏，或者说减轻工艺人员的脑力劳动，另一方面，可以避免单个专家的片面性、局限性和一时疏忽，结论比单独专家更加准确。

4.6 本章小结

本章主要介绍了系统的模块及运行实例，包括用户界面设计、登陆模块、查询模块、工艺设计模块、以及数据挖掘模块的具体内容。其中，作为课题创新点的数据挖掘模块的设计，本章举出测试实例来进行运行实证，解释了焊接工艺潜在规则提取的具体方法及挖掘结果所包含的意义。

第五章 研究结论与创新点

本文对互联网上以及焊接生产中大量离线的工艺数据进行深入的研究,结合数据挖掘技术,对焊接生产工艺数据特点进行分析,设计了基于 Apriori 算法的关联规则改进算法,并应用数据挖掘技术开发了创成式焊接 CAPP 系统。经过理论分析,计算机建模,软件编制以及试验工作,得出如下结论:

(1) 课题在分析已有焊接 CAPP 系统的不足和焊接工艺设计的特点的基础上,结合工厂的实际生产情况,提出了基于数据挖掘技术的焊接 CAPP 系统设计思想,确定了系统的总体框架及各模块的功能。

(2) 选择具有强大数据库功能的可视化编程工具 Delphi7.0 进行焊接 CAPP 系统界面设计、功能实现和工艺数据的管理,界面简洁友好、图文并茂、方便直观,便于系统的管理和使用。

(3) 利用 Delphi 的外挂数据库 Database Desktop,建立了焊接资料库,如母材成分库、母材组织性能库、母材力学性能库、焊材成分库、焊材组织性能库、焊材力学性能库、保护气体特性用途库等,为系统提供了标准规范的格式和比较完备的管理和查询工具,方便了用户的查询。

(4) 采用关联规则改进算法,在数据库管理系统中引入数据挖掘技术作为其智能工具组件,实现焊接工艺的制定。用户只需向系统调入焊接工艺文件,系统将自动进行潜在规则的提取和挖掘过程。

本文的创新之处在于:

(1) 与经典的 Apriori 算法相比,基于 SQL 语言的关联规则挖掘算法可以直接对数据库中的数据进行操作,只利用 SQL 提供的 9 个动词即可满足用户从建库到查询、数据维护和简单统计等对数据库的操作需求,无须再用烦琐的程序来控制循环,具有较高的执行效率和较低的算法复杂度。

(2) 根据创成式 CAPP 系统构建的基本原理,改进为基于数据挖掘技术的焊接 CAPP 系统,实现了数据挖掘模块功能,对焊接生产中工艺优化有较高的实用性。

(3) 本文根据焊接工艺数据库系统的现状和数据挖掘技术特点,开发了基于数据挖掘技术的创成式焊接 CAPP 系统,结合了焊接工艺设计和数据挖掘技术两个不同的学科,是焊接工艺设计和管理计算机一体化的新需求。

参考文献

- [1] 李培根, 张国军, 关于 CAPP 的实践与思考[J], 中国工程科学, 2005, 7 (3): 13~16
- [2] 董黎敏, 朱世和, 徐燕申等, CAPP 的发展及研究现状[J], 天津理工学院学报, 2004, 20 (2): 43~45
- [3] 恒永兴, 张振明, 孔宪光等, CAPPFramework 设计与实现关键技术研究[J], 计算机工程与应用, 2002, (10): 116~118
- [4] 戴同, CAD/CAPP/CAM 基本教程, 机械工业出版社, 1997: 55~57
- [5] 宁汝新, 赵汝嘉, 欧宗瑛, CAD/CAM 技术, 机械工业出版社, 2005: 145~153
- [6] 彭梨花, 樊瑜瑾, 李浙昆等, CAPP 技术发展状况与趋势[J], 昆明理工大学学报, 2003, 28 (5): 68~71
- [7] Marri H B, Gunasekaran A, Grieve R J. Computer-aided process planning: a state of art [J], International Journal of Advanced Manufacturing technology, 1998, 14 (4): 261-268
- [8] 陈宗舜, 论 CAPP 的发展方向[J], CAD/CAM 与制造业信息化, 2003, (11): 8-11
- [9] 乔尚飞, 魏艳红, 计算机辅助焊接工艺设计应用现状[J], 焊接, 2002(7): 5-8
- [10] 王玉玲, 刘长安, 周以齐, 焊接 CAPP 系统的研制与开发[J], 机械工艺师, 1999 (9): 12-13
- [11] Aldakhilallah K A, Ramesh R. An intergraded framework for automated process planning [J], Design and Analysis, 1998, 36 (4): 939-956
- [12] 朱志明, 张崇轲, 陈丙森, 焊接结构件装焊 CAPP 系统的研究与开发[J], 焊接学报, 2001, 22 (4): 87~91
- [13] 杨志刚, 王先逵, 刘成颖等, 动态联盟环境下的辅助工艺规划系统, 清华大学学报, 2001, 41 (2): 46—49
- [14] Tu Yi-liu, Chu Xu-lin, Yang wen-yu, Computer-aided process planning in virtual one-of-a-kind production[J], Computers in Industry, 2000, 41 (1): 99—110
- [15] 吴伏家, 王感苍, CAPP 的研究和发展[J], 华北工学院学报, 2001, (22): 6
- [16] 黄乃康, CAPP 技术发展现状与趋势[J], 计算机辅助设计与制造, 2001, 5: 3—5
- [17] 王正刚, CAPP 系统工程化的关键技术研究[J], 机械设计与制造工程,

- 2002, 31 (2) : 65-67
- [18] 陈永府, 王峰, 朱林等, CAPP 发展趋势及面临的问题[J], 计算机工程与设计, 2004, 25 (5) : 776-781
- [19] Vanden Berghen Frank. Classification trees: C4.5[EB/OL], <http://iridia.ulb.ac.be/~fvandenb/work/classifier/classifier.pdf>, IRIDIA, University Libre de Bruxelles, 2003
- [20] Chidanad Apte, sholom Weiss. Data mining with decision trees and decision rules[J], Future Generation Computer Systems, 1997, 13: 197~210
- [21] 章兢, 张小刚, 数据挖掘算法及其工程应用[M], 机械工业出版社, 2006: 3~4
- [22] 赵丹群, 数据挖掘原理、方法及其应用[J], 图书馆自动化, 2000, (6) : 41~44
- [23] 田艳, 数据挖掘技术的应用与发展[J], 统计与信息论坛, 2004, 19(4) : 18~21
- [24] 冯兴杰, 周淳, Apriori 算法的改进 [J], 计算机工程, 2005 (7), 172~173
- [25] 王伟勤, 钟敬堂, 对 Apriori 算法的一种改进 [J], 佛山科学技术学院学报, 2007, 25 (2) : 54~57
- [26] 柴华昕, 王勇, Apriori 挖掘频繁项目集算法的改进 [J], 计算机工程与应用, 2007, 43 (24) : 158~162
- [27] Quinlan J R, Induction of Decision Tree [J], Machine Learning, 1986
- [28] 王迎春, 基于决策树算法 C4.5 的冲压工艺知识发现[J], 机械科学与技术, 2004, 23 (12) : 1506~1514
- [29] Quinlan J R, C4.5: Programs for Machine Learning [M], Morgan Kaufman, 1993
- [30] 郑金桥, 蒋向华, 王义林等, 汽车覆盖件冲压 CAPP 系统的研究 [J], 模具工业, 2005 (3) : 7~11
- [31] 张文芳, 韩晓健, 基于约束的创成式 CAPP 系统研究与开发[J], 新技术新工艺, 2006 (4) : 21~22
- [32] 林三宝, 黎明, 吴林等, 基于特征和面向对象的焊接结构产品信息模型研究[J], 焊接学报, 1999, 12: 138~145
- [33] 姚宗湘, 计算机辅助焊接工艺设计系统的开发与研究[D], 重庆大学硕士学位论文, 2005.3
- [34] 蔡建刚, 基于专家系统的焊接生产定额计算机辅助制订系统的研究与开发[D], 兰州理工大学硕士学位论文, 2006.6

- [35] 陆惠恩, 陆培恩, 软件工程简明教程[M], 电子工业出版社, 2005: 26~36
- [36] 胡玉忠, 钢制压力容器焊接 CAPP 技术及实现[D], 合肥工业大学, 硕士学位论文, 2003.3
- [37] 邵国良, 宗培, 曹雷, 焊接施工报告数据库管理系统设计[J], 焊接技术, 2005, 34 (3) : 56~58
- [38] 赖炼, 谢娜, 王春伟, 焊接工艺设计数据库软件系统[J], 焊接技术, 2005, 34 (4) : 22~24

硕士期间发表的论文

- [1] 钱学君, 胡小建. 决策树算法在焊接工艺设计中的应用研究[J].焊接技术, (已录用).
- [2] 钱学君, 胡小建. 焊接 CAPP 系统的应用概况及前景展望[J].现代焊接, 2009, 9: 33~35.