

摘 要

近年来，随着钢铁工业的竞争的加剧，降低能耗，提高产量和质量成为球团厂生存和发展的必由之路。在球团生产自动化系统中，调度数据库服务器中就保留了大量的生产数据，合理有效地运用这些生产数据可以优化调度提高效益。本论文的任务就是利用现有的生产数据实现熟球产量的预测与合理得生产计划，在保证产量的前提下节约原料。

本文首先分析了流程工业生产调度的特点和现存问题，回顾了生产调度研究的发展和现状，在此基础上结合球团厂现有生产调度的缺陷和生产数据的特点，并根据球团工艺，通过聚类统计和相关性分析提出了数据挖掘的对象即产量与原料用量和质量间的关系。随后，考虑到单纯的产量并不体现实际的生产状况，提出创建产量与质量相结合的综合产量属性作为预测目标。而通过对球团厂长期的产品综合质量的评估，得出产品质量稳定的结论，避免了创建新属性。然后利用模糊神经网络挖掘产量预测模型，给出产量与原料质量和用量之间的对应关系模型，实现了熟球产量的预测，并在此基础上，改变网络输入输出，得出生产计划模型，根据一定的计划产量合理安排各原料用量。

本文针对大冶球团厂生产数据的特点在众多挖掘技术中选择了统计类数据挖掘技术，采用回归和最小二乘法给出产量与原料用量的线性关系。在分析了该线性模型的不足的基础上，利用决策树挖掘技术，加入了原料质量指标中对产量影响相对较大的属性，完善了产量预测挖掘模型。在无法确定模型中属性的非线性关系的情况下，引入神经网络技术，利用反向传播神经网络的学习能力来训练模型。在最终的模型表达中运用模糊理论，避免了神经网络的过度拟合和不稳定性，加强了模型的适应性和可靠性。

关键词：数据挖掘； 流程工业； 生产调度； 模糊神经网络

Abstract

Nowadays, with the competition of the iron and steel industry, it's the only way for the Pellet Factory's surviving and development to reducing the energy consuming, increasing the output and the qualities. Lots of production data basing on the time sequence remain in the control center database of the production automatic system. Using these data reasonably and efficiently can optimize the scheduling and increase the profits. This thesis is going to utilize these existing data, in order to forecast the output and distribute various raw materials according to the proposing production.

First, the thesis analyzed the specialties and the existing problems of the controlment of the flow industry producing, and then reviewed the developments and the actual state of the research on the subject of production direction. Based on these above and in accordance with the ongoing systems' bugs and the production data's characteristics, through clustering statistics and relativity analysis, a statistics model was brought forward, then I present the model for production prediction. Secondly, during the construction of this scheme model, at first the evaluate of the products' integrated qualities in a long range resulted that the quality is stable. Following, with the usage of the fuzzy neural network the production prediction model was consummate. Finally, based on the production prediction model a production planning model was brought out .

The thesis aimed at the specialties of the production data and so the techniques of statistics datamining was choosed. Adopting the Least Square Estimation Method, the statistics model of the relationship between the inputs and outputs was given. As the validation of this model was not good, the neural network techniques was introduced and the BP neural network was used to achieve the nonlinear model. In the final expression of the rules the fuzzy theory was introduced and it successfully avoids the over fitting and the instability of the neural network, so it consolidates the model's flexibility and reliability.

Keywords: Datamining; Flow Industry; Production Direction; Fuzzy Neural Network

独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：徐晓瑾

日期：2004年4月30日

学位论文授权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权华中科技大学可以将本学位论文的全部或部分内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和汇编本学位论文。

保密 ， 在 _____ 年解密后适用本授权书。

本论文属于

不保密 。

(请在以上方框内打“√”)

学位论文作者签名：徐晓瑾

日期：2004年5月1日

指导教师签名：李可富

日期：2004年5月1日

1 绪论

数据挖掘是从大型数据库或数据仓库中发现并提取隐藏在其中的信息的一种新技术,它能自动分析数据,进行归纳性推理,从中发掘出潜在的模式。数据挖掘技术涉及数据库、人工智能、统计分析等多种技术^[1]。一个大中型工业企业,一般都有大量与生产设备运行相关的数据,生产调度更是一个基于知识的决策过程,因此数据挖掘技术应该在生产调度中有用武之地。本文主要研究数据挖掘技术在球团生产调度中的应用。球团生产是典型的流程工业,流程工业生产调度负责将计划分解并下达给各个生产装置,同时还要收集、处理生产现场数据,及时地响应各种突发事件以保证整个生产工艺流程协调、均衡地运转。球团生产调度系统中保存了每日的产耗统计及质量分析报表,通过对这些数据的挖掘利用,可以指导生产计划。

统计类数据挖掘工具可以在时序数据和序列数据中发挥重要作用,主要是趋势分析、相似性探索、与时间有关数据的序列模式挖掘和周期性模式的挖掘^[2]。根据球团生产数据的特点和本论文的任务即利用现有的生产数据实现熟球产量的预测,调整短期生产计划,我们选用统计类数据挖掘技术。

本章简要介绍了球团生产调度系统研究发展和现状、课题的来源和意义以及本论文研究的重点内容。

1.1 课题背景

本课题是以大冶球团厂的生产自动化系统为背景的。球团厂的主要产品是熟球,主要提供给武钢作为炼钢的原料,其生产状况不仅决定了自身的发展还影响了武钢的生产和发展。

迄今为止,我国钢铁工业一直是依靠国家政策的保护发展起来的。就是国家实行进口许可证制度和用征收关税的办法,控制国外廉价的钢铁产品进入国内市场,从而保护了发展中的我国钢铁工业,使之在少受外来干扰的环境中依托国内市场需求的增长而迅速发展起来。但是,在国际贸易中,要讲对等互惠原则,限制别国的产品进口,自己的产品出口别国也将受到限制。随着我国经济发展到现在的水平,继续设置贸易壁垒已经弊大于利,加入世界贸易组织已成大势所趋。参加世贸组织

华中科技大学硕士学位论文

就意味着国外产品要享受国内待遇，就是和本国产品一样的待遇，所有进口商品的关税国际上一般只是 3%~5%，而且像许可证等非关税壁垒都要取消，就是说要搞自由买卖。到那时候，中国的钢铁工业就要真正面对国际市场。从受到贸易保护到参与国际市场竞争，这对于我国钢铁工业来说是一个重大的转变。面对国际市场，中国钢铁工业怎么办？^[3]

目前我国钢铁生产技术经济指标落后，有以下几点：

- 1) 能耗。我国钢铁生产能耗高，不管是炼铁焦比和其它工序能耗都远高于国际先进水平，结果我国产 1 吨钢的综合能耗达到 1.5 吨标准煤，而在日本，生产 1 吨钢的综合能耗还不到 700kg 标准煤，就是说中国产 1 吨钢所用的能源相当于日本的 2 倍多，所以这项指标相当落后。
- 2) 物耗。我国钢铁生产各工序材料消耗高、钢的成材率低。像日本，由于推进了全连铸新工艺，全国平均成材率已经达到 95%，而我国，武钢算比较好的，才 86.7%，全国重点企业平均计算大概是 82%~83%，和日本比差距很大。日本 1973 年最高产钢 1.2 亿吨，当时钢的成材率是 80% 以上，那时候日本生产的钢材还没有它现在产钢 9000 万吨时多。现在多生产的钢材都是靠提高成材率得来的。可见，提高成材率对增产钢材是有很大的潜力的。
- 3) 劳动生产率。80 年代末，日本的钢铁厂全员劳动生产率一般是 500 吨钢 / 人·年，欧洲很多厂也是 400~500 吨钢 / 人·年，中国平均是 20 多吨钢 / 人·年。武钢 12 万人，产钢 500 万吨，人均劳动生产率是 40 多吨钢 / 人·年，比全国平均水平高。宝钢全员是 3 万多人，去年年产钢 600 多万吨，人均 200 多吨钢，是全国最高的。实际上，劳动生产率的差距是我们和西方国家差距之中最大的。
- 4) 设备寿命、作业率。我们现在的设备寿命、作业率也比别人低。以高炉为例，我国高炉一般每 2~3 年就要停炉中修，而国际上高炉炉役的标准水平是 80 年代连续生产 8~10 年，现在是 15~20 年。日本一座高炉已连续生产了 17 年，现在还在继续生产。我们在这方面的差距也是很明显的^[3]。

面对钢铁工业中存在的以上这些缺陷，站在球团工业的角度，可以从提高成品

产量和质量,提高原料利用率等几个方面改进。

而目前我国球团生产企业中,对生产过程的实时监控大多依然局限在车间控制室一级,厂级的生产调度和操作管理仍然停留在通过电话传递信息及手工处理报表的初级阶段,为了更有效的指导生产,应该建构一个综合决策、生产调度、车间管理、过程优化为一体的计算机集成生产自动化系统。

本研究课题来源于武钢矿业有限责任公司大冶铁矿竖炉球团生产过程自动控制系统。竖炉球团生产是一个集制气、配料、烘干、润磨、造球、布料、烧结、出料等于一体的复杂过程。从现有生产过程来看,大冶竖炉球团存在一些值得改进的地方。随着用户对球团质量要求的不断提高及国内外竞争的加剧,传统的人工控制竖炉球团生产过程的方法已经不能适应生产发展的要求。所以,充分利用现代计算机技术、网络技术、检测技术及现代控制技术,实现竖炉球团生产过程的自动控制、自动监测和最优调度,这是实现球团生产面对市场竞争的唯一出路。

1.2 课题目的及意义

本课题根据大冶球团厂及其所属矿业公司的生产状况和发展需求,以球团厂生产综合自动化系统为基础,对生产调度数据库中的数据进行挖掘,通过合理有效地利用这些生产数据和人工统计数据,建立基于产耗的生产调度系统,实现产量的预测,指导生产计划的制定,有效地进行短期计划调整。

如何从这些大量数据中提取出隐藏在其中的有用信息的,来进行有效的生产优化调度是本文研究的主要问题。

无论是查询、统计还是报表,其处理方式都是对指定的数据进行简单的数字处理,而不能对这些数据所包含的内部信息进行提取。企业要提高竞争力,需要大量有价值的信息来辅助决策,单单是查看一些长期累积下来的数据,并不能给企业生产提供帮助,必须对数据进行更高层次的分析。

数据挖掘在生产调度中的应用不但有必要,也有成功的可能:

1) 生产调度是一个决策的过程 正确的决策建立在对知识的掌握程度上。掌握的知识越完整、越丰富、越可靠,决策的正确性就越大。它需要以集成数据为基础,但现实中的数把往往是分散管理的且大多分布于异构的数据平台,数据集成不易;

就一直试图使这一得力工具具有更多的智能，替代更多的人力。在生产调度领域，人们希望计算机能够：

- 1) 模仿调度专家，应用已有的经验和知识；
- 2) 模仿调度专家，学习已有的知识；
- 3) 模仿调度专家，发现潜在的、未知的规律。

一般说来具有以下特点领域都可以采用数据挖掘：

- 1) 有大量充足的相关数据；
- 2) 需要基于知识的决策；

在大中型企业，一般都有大量与生产设备运行相关的数据，生产调度更是一个基于知识的决策过程，因此数据挖掘技术应该在生产调度中有所用武之地。

但是与零售业、金融业等领域的信息相比，生产调度信息有以下特点：

- 1) 多样性。生产调度信息的多样性体现在以下方面：
 - 信息来源多样。生产调度信息的来源很多。如它可以来自各类生产设备，也可来自不同的生产部门，还可以来自不同的地域，这些都使生产调度信息多样化；
 - 数据类型多样。与商业、金融业等不同，生产调度用到的数据类型不但有常规的整型、实型、布尔型，还有文本等类型；
 - 数据格式多样。在生产调度领域，由于还没有一个统一的数据存储和交换标准，使得数据格式各不相同。

- 2) 复杂性。

调度信息的复杂性一方面是由于它的多样性，另一方面是由于信息的成分也十分复杂。在调度信息中，最多的可以说是生产设备数据，生产设备数据的一部分来源于各类 DCS 的生产过程中实时采集，一部分来源于生产相关人员的手工输入，还有来源于调度系统的自动生成。这些数据都不可避免地带有噪声，但这些噪声地来源和特性是不一样地。它们与有用信息地叠加不仅是线性叠加，还有非线性叠加。

- 3) 隐蔽性。

有的调度规则可由相应的调度数据中获得，但更多情况下，调度数据与调

度规则间没有必然对应关系。这时需要数据变换和信息集成与融合。数据变换是从不同角度来发现调度规则，如数量特征不明显的规则可能在时间特征方面一目了然；信息集成与融合是通过增加信息量使隐蔽的调度知识显示出来。调度知识的隐蔽性使生产调度中始终面临的一个问题，多表关联挖掘就是充分利用信息的一个典型例子。

此外，生产调度还有自己完整的理论和方法，并且与工程实际有着紧密的联系。因此，数据挖掘要在生产调度中获得成功应用，专业领域知识是必不可少的。

可见，挖掘对象的复杂性和专业领域知识的特殊性使影响数据挖掘在生产调度领域成功应用的主要因素。

1.4 本文主要研究内容

本文针对球团生产数据的特点，对生产数据库进行数据挖掘从而优化生产调度。主要研究内容有：

- 1) 流程工业生产调度的特点和方案；
- 2) 产量预测模型的线性回归挖掘；
- 3) 挖掘原料质量中对产量的影响因素，完善产量预测模型；
- 4) 利用模糊—神经网络挖掘产量预测模型和生产计划模型。

2 流程工业生产调度

本文是将数据挖掘技术应用于流程工业生产调度系统中，所以我们首先从分析流程工业生产调度的特点入手，进而选择合适的生产调度方案，确定数据挖掘对象和目的。

2.1 流程工业生产调度特点

流程工业企业的生产过程具有如下的特点：

- 1) 原料处理量大，生产过程可以连续不间断；
- 2) 产品品种相对稳定，工艺流程基本不变，但工艺参数多变；
- 3) 生产控制实时性要求高，因而普遍采用了集散控制系统（DCS）；
- 4) 生产装置安全、稳定、长周期、满负荷、优质运行是企业实现低成本、高利润的关键。

流程工业的生产过程的这种连续性、产品稳定性、生产量大等特性、其产品常常不是以新取胜，而是以质量和价格取胜，因此，流程工业生产调度与离散工业生产调度在具体的实现上有很大的不同，它具有如下的特征：

- 1) 良性结构问题与不良结构问题并存。流程工业生产调度问题本身是不良结构问题，但其中包含有良性结构问题。例如生产调度中的作业计划制定是一个规划问题，可以建立线性规划模型，用线性优化方法求解。而对于实时调度问题，用纯粹的数学方法很难求解，必须引入人工智能的理论与方法进行求解。
- 2) 实时性。从根本上说，任何生产调度都有实时性的要求。对于流程工业生产调度，出于生产是在连续不断的进行之中，调度问题也随着生产流程的变化而变化，在时间上要求调度决策迅速及时，与生产流程保持同步，要求滞后时间在一定的域值范围之内。
- 3) 全局调度与局部控制相结合。作为指挥全厂生产的生产调度管理，必须在

宏观上把握生产的全过程，又必须对每一个车间和装置的具体生产做出指导，以保证生产高效、高质。

- 4) 处理突发事件，保证生产的平稳进行。在实际的生产中，存在许多非常规的生产情况。例如某项生产控制指标超出临界值等。要求生产调度能监控生产情况，分析情况，及时做出处理决策，保证生产的正常运行。

2.2 生产调度的方案

习惯上，生产企业可按过程特征分成连续型、批量间歇型和离散型三类。这三类企业的控制对象具有不同特性，因此数学描述不同，如连续企业多采用微分/差分方程，离散企业则采用离散事件动态驱动系统（DEDS）模型。以计算机集成制造（CIMS）的现点来看，各种企业都呈现递阶层次结构，除对底层对象所采用的控制方有所差异外，生产调度、经营决策等层次所体现的特征和实现的功能是一致的，也就是说不会因为企业生产过程特征的不同而分门别类地研究生产调度。因此，可以从结构共性来研究不同企业的生产调度问题。

生产调度是指“产品在制进过程中，在满足各种生产约束（如作业优先级、设备能力、交付日期等）的前题下，实现人力、材料、机器等共享资源的有效配置及使用顺序，以达到生产费用最低的目的。简单地说，生产调度就是在一定的时间范围内为完成特定的生产任务而分配共享资源，并使得预定的某些生产指标最优。

人们之所以对生产调度如此感兴趣，而且不遗余力地去研究它，最根本的一个原因是一个合理的调度方案能给企业带来很大的经济效益。另外在学术上还因为它存在以下特点：

- 1) 从数学角度讲，调度是一个多目标、多约束的优化问题，是一个 NP 完全问题（Non-polynomial complete）；
- 2) 几乎每一个生产环境都是唯一的，很难用一个生产环境的调度方案去解决另一个生产环境的生产调度；
- 3) 调度领域知识的多样性；
- 4) 生产环境的动态性。

生产调度的以上特点，使得生产调度的研究具有很大的复杂性，各种调度技术

都有其适用条件和局限性，不存在一种通用的解决办法。目前，求解调度的方案主要可分为两类：一类是基于模型的生产调度，另一类是基于规则的生产调度。本文采用基于规则的生产调度。

2.2.1 基于模型的生产调度

这种方法的求解步骤是：首先在定义调度问题的基础上建立调度模型（常用的有数学规划模型、控制系统模型等），然后基于该模型运用一定的调度算法进行求解。这类方法的特点是根据一定的性能指标进行寻优，对规模小、相对简单的问题能得到令人满意的全局最优解，但随着问题规模的增大，求最优解将变得不可行。

1) 数学规划模型

又称资源规划，它研究的是一种最简化的形式：对 N 种产品在 M 个生产装置完成各自工艺而对共享资源进行优化达到所需目标的最优。这种对 $N \times M$ 资源共享问题的研究揭示了调度问题的组合特性引起的计算复杂性。它主要用来讨论静态调度。常用得数学工具有：排队网络方法，分枝定界方法，混合整数线性规划（MILP）模型，混合整数非线性规划（MINLP）模型，多目标决策等。这些模型的求解算法在时间花费上大都随规模的增大而呈指数增长，但是数学规划仍然是静态生产调度的主要数学依据。

为了解决大规模数学规划带出的求解复杂问题，最近几年，人们在数学规划的求解过程中引入了人工智能的方法，如递归神经网络方法，遗传算法，模拟退火算法，启发式问题求解方论。这几种方法都是人工智能领域中重要的搜索算法，这一趋势将大力推动静态调度数学规划求解的发展。

2) 控制理论模型

利用控制理论来研究生产调度，研究对象主要是动态调度问题。它把调度系统看成一个动态系统：资源数量、产品需求、生产能力及装置故障信息等构成系统输入，产品在制品数量、装置状态（闲置、修理、准备、正在加工）等构成系统状态，系统状态的组合方式为控制系统的输出。这一理论的一个优点就是考虑了机器故障、加工失败等意外事件的发生。这一方向有可能会有重大突破。

2.2.2 基于规则的生产调度

这里的规则不仅包括调度规则，而且包括利用调度规则的规则和经验。在生产调度中应用规则有两种方式，一种是在上面讨论的基于数学模型的方法中，利用启发式方法求解。这种方法是利用领域知识，引入调度规则，减小问题的求解空间，进而大大减少了较大规模数学问题求解所花费的时间，并且能解决一定的动态调度问题，代价是解得的是一个可行解，而不是一个最优解；另一种方式就是应用于本部分将要介绍的基于规则（而非数学模型）的生产调度中。

这种方法完全根据一定的规则或策略来确定生产系统中的下一步操作。通过应用各种调度规则，能够为局限于规则覆盖范围的问题产生合理的解决方案。这种方法的特点是不必进行大量的计算，有了合适的规则以后可以很快地生成调度方案；另一优点是使动态调度问题得到了解决而动态调度是数学规划方法的一个死角。

最早的基于规则的生产调度是人类专家利用经验知识、大量的过程基础知识和丰富的联想能力，处理生产调度环境中出现的突发事件和其他一些不确定性问题。但出于作为专家的人的计算能力和记忆能力有限，当问题规模较大时，也是很难解决的。

近几年，人工智能同样被引入基于规则的生产调度，用它结合人类调度专家的经验，对生产调度问题进行建模并求解。这里的建模不是指建立数学模型，而是指用人工智能中的知识表达技术表示领域内的有效知识。目前常用的方法有：应用产生式规则的调度，应用神经网络的调度，应用 Petri 网的调度以及智能决策支持的调度。

基于规则的生产调度方法得到越来越多应用的同时，随之而来的问题是如何快速选用合适的调度规则以及如何挖掘出一些新情况下的新规则。

1) 应用产生式规则的调度 (Production Rule Scheduling)

这种方法是把调度规则作为专家系统的知识，经过推理（一般采用模糊推理）产生启发式结果提供给调度人员，由于产生式规则是由有经验的决策管理人员和工程技术人员总结出的一套行之有效的经验法则，是启发式的，非常符合人们利用经验处理问题，所以目前它是最常用的知识表达和建模方法。在这方面 Kanet et al. 等研究了基于规则的专家系统。这一方法的优点是简单、实用，花费的求解时间少。

但当系统规模比较大时，由于调度规则本身就涉及到大量数据的检索、排序、以及生产过程状态的判断等，使得规则表示的前提部分变得比较庞大，失去了规则表示法简单明了的特征，造成规则匹配速度慢；另外这种方法自适应能力（学习能力）差，无法适应动态的调度环境。为了解决匹配速度慢的问题，一些人适当对规则进行分类，这样可以更清晰地表达经验知识，进而加快规则匹配速度。但自适应能力差的问题仍无法解决。

2) 应用神经网络的调度

由于通过训练学习可以使神经网络连结权上存贮大量领域知识，即神经网络具有学习能力，所以神经网络被广泛应用于知识的表达和规则模型的建立方面。但利用神经网络进行建模并进行调度求解还存在以下问题：

神经网络的知识表达和学习训练是一对矛盾，神经元个数较少时，虽然训练较为容易，但知识表达能力太差，网络大了，表达知识强了，但训练学习起来又难了。所以一般合理设计神经元的有限个数，既能保证有效知识的表达，在生习训练时又不至于过于困难。

神经网络的学习都是基于现有知识的基础之上的，学习的结果也就只是记忆了在现有知识下的一种输入/输出映射或一种存储模式。这样若当前状态位于已学知识之外，则现有的神经网络很难得到与当前状态相关的正确结果。所以，人们对神经网络的研究主要集中在较快和较为有效的学习方法和有利于再学习的网络结构上。

另一个问题是由神经网络的天生缺陷造成的，即网络权值意义不明确。它只能描述难以理解的大量数据之间复杂的函数关系，无法以人类语言可以接受的方式表示出来。一旦学习结束，网络的结构和权值也就确定了，如要调整输入和输出的对应关系，不能简单地调整一个或几个全值，只能重新学习、集体调整。为了克服这个缺点，本论文采用了一种基于神经网络的模糊推理系统^[4]。

总之，用神经网络建立模型求解调度主要要解决的问题有，学习速度，调整知识更清楚表达和规律再学习。本论文的任务就是利用现有的生产数据在统计挖掘的基础上运用合理的推理系统和合适的规则获取系统构成基于产耗的生产调度系统，模糊推理系统的分散式结构使得再学习变得容易，通过选用较快的学习算法提高推理和学习速度。

3) 应用 Petri 网的调度

Petri 网作为一种图形和数学工具,为离散事件系统的模型化,性能分析和设计提供了一个统一的环境。Petri 网是一种特殊的双枝有向图,用位态(place)表示系统的状态或条件,变迁表示系统的事件,有向弧表示状态和事件之间的联系,因此用 Petri 网模型化一个过程较为直观易懂,缺点就是由于不同的变迁具有不同时开放规则,所以很难用一种传统的数学分析方法去分析它的状态演变:当系统规模较大且较为复杂时,Petri 网的规模也大大增加,如此,便失去了其原有的直观性。为了解决这一问题,对 Petri 网的分解简约方法的研究越来越受到人们的重视^[5]。

4) 决策支持的生产调度

用人工智能的方法解决调度问题,关键在于如何模仿调度专家处理调度问题的能力。虽然上面提到的基于模糊神经网络的生产调度系统能解决一些数学规划不能解决的调度问题,但有时出现的调度问题单纯依靠这种人工省能的方法是解决不了的或解决不好的。原因在于知识库存贮的知识较人类来讲,面比较窄,借鉴能力相对较差;另一方面虽然说计算机处理数据速度已大大提高,但就某些定性推理和认知方面还较人类差的远,所以人工智能的利用再加以人类调度专家的适当指导,将会大大加强调度问题的求解能力,更好的适应生产调度环境的变化,在本课题中生产调度系统采用的框架格和人工智能结合人类调度专家指导的智能决策支持系统。这种智能决策支持系统的理论基础就是神经网络的规则推理和挖掘^[6]。

2.3 球团生产调度的过程描述

流程工业的生产过程具有流程连续、物流不能间断、产品相对固定,生产操作、过程监控要求实时性强等特点,企业生产以均衡物料、均衡能源、安全高效、低耗优质、优化工艺操作为目标,工程特征大都表现为大范围不确定、高度非线性、强关联性和交错的信息结构^[7]。

球团工业是典型的流程工业,建构一个综合决策、生产调度、车间管理、过程优化为一体的计算机集成生产自动化系统,是当前流程工业发展的方向。

而目前球团厂对生产过程的实时监控大多依然局限在车间控制室一级,厂级的生产调度和操作管理仍然停留在通过电话传递信息及手工处理报表的初级阶段。

对这种类型的企业来说，要实现 CIMS，就要运用 CIMS 思想，在不同时空层次上，实现物质流和信息流的综合控制和管理，构成一个综合决策、生产调度、车间管理、过程优化为一体的计算机集成生产自动化系统，以球团工业为例，其集成框架如图所示。

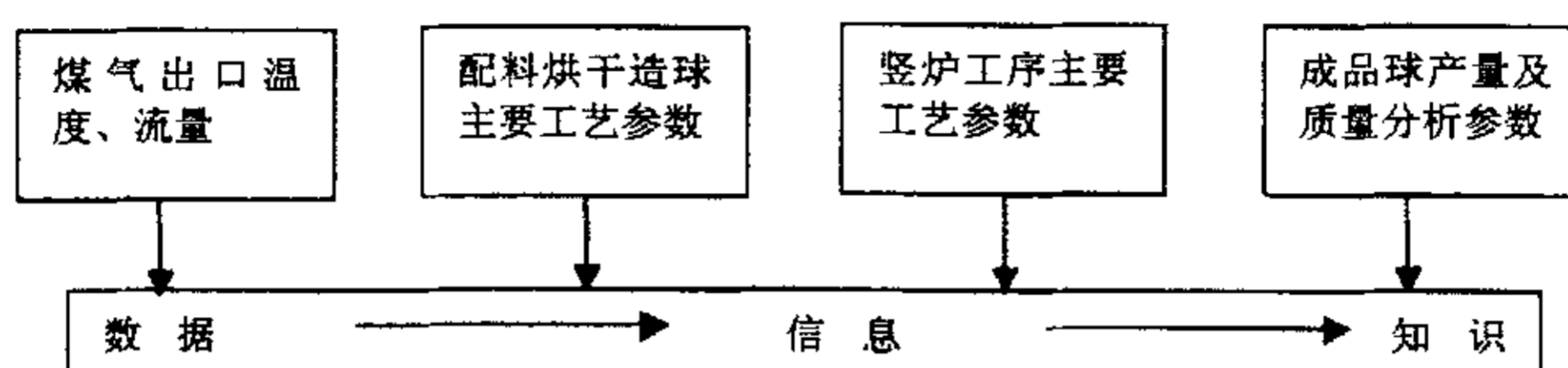


图 2-1 球团生产信息集成框架

由图可知球团生产过程的信息集成是一种多层次、结构错综复杂的系统，为了实现球团企业的 CIMS 总体设计，针对上述的球团企业生产信息集成，可将流程工业 CIMS 分为三个层次：最优计划决策、最优调度和最优控制。

控制层是分秒级实时运行的，优化层的最优化计算，根据工艺的实际情况，对于优化调度是以天或旬计算一次，而计划是按月、季或年进行的。应该说，在最优控制、最优调度、最优计划决策所有这些层次上都存在优化问题，都可分别获得不同的经济效益。最优控制，它是在线寻求过程的操作条件最优，使生产过程处于最优工况附近，从而获得经济效益；最优调度，它是借助合理调度一个大型企业的原料、中间产品、能源以及恰当地处理在连续流程工业过程中出现的故障，保持均衡生产等以寻求最大的经济效益；长期计划决策是根据市场情况、订货合同来最佳的组织生产，减少流动资金、减少库存、最优排产等来获取最大限度的经济效益。

生产调度是完成计划（年度、季度、月）的具体措施和保证，从而保证全年利税指标得以完成。同时，生产调度协调各生产装置按照生产流程的顺序最佳地分配原材料、能源，使全厂生产效益最好。调度调优的目标可归纳为在给定生产计划任务、设备能力、人力、指标的前提下，实时组织协调工艺参数，使各生产装置均达到最优（局部最优），同时还必须保证企业全局最优，是流程 CIMS 的关键技术之一。

球团生产调度包括三个阶段，分别是作业计划制定，现场生产调度及调度后处理。下面分别进行描述：

1) 生产作业计划制定。生产计划制定是根据生产管理层下达的月计划生产指

标, 例如生球、熟球产量, 成品球质量指标等, 对月计划进行分解, 考虑全局优化和局部控制, 制定每日作业计划, 即具体到每日熟球产量及对应的工艺参数指标。

- 2) 生产现场调度主要是根据日作业计划下达生产指令给各车间与装置, 指导全厂生产, 并根据反馈的各项生产指标, 及时调整生产计划和生产方案, 并处理生产中的突发事件。
- 3) 调度后处理的主要任务是根据日作业计划和生产状况进行各工段生产工艺参数优化, 制定中间产品的技术指标 (如生球水分, 煤气质量等等)。

文中, 根据球团调度现状, 将调度系统的实现重点放在作业计划制定和现场生产调度这两个阶段。

2.4 球团生产调度方案设计

根据以上的球团生产调度的过程描述, 并结合球团厂生产现状和现有的生产数据库, 本文采用基于规则的生产调度方案。球团调度的重点在作业计划制定和现场生产调度这两个阶段, 因此本文把所要挖掘的规则定为熟球产量与各原料质量和用量之间的对应关系。

为规则简化而更具适用性, 在最终的规则表达中运用模糊理论, 将各质量指标属性由数据值型转换为枚举型的质量等级属性, 运用合理的推理系统和合适的规则获取知识构成基于产耗的生产调度系统, 实现熟球产量的预测和生产计划的制定, 在保证产量的前提下节约原料。

2.5 本章小结

本章在讨论了流程工业生产调度特点的基础上, 分析了现有的求解调度的方案及其缺点, 并对球团生产调度现状和调度重点做出分析, 确定了球团调度的方案和数据挖掘的对象和目的。下章将对产量预测模型的挖掘作详细的阐述。

3 产量预测模型

在球团生产调度系统中, 调度服务器数据库中就保留了大量的时间序列的生产数据和每日的产耗统计数据及质量分析数据, 最初, 我们利用这些数据制定生产报表, 拟合主要生产参数变化的曲线。而这些都是不能为生产调度提供直接的决策支持。需要从这大量数据中提取出隐藏在其中的有用信息, 来发现现有数据中的关系和规则, 根据现有数据预测未来的发展趋势从而采取相应的行动, 为此, 我们采用了数据挖掘技术。

统计类数据挖掘工具可以在时序数据和序列数据中发挥重要作用, 主要是趋势分析、相似性探索、与时间有关数据的序列模式挖掘和周期性模式的挖掘^[2]。

本章结合球团生产工艺过程, 通过聚类分析和相关性统计分析提出了数据挖掘的模型: 铁精矿用量、铁精矿硫份和焦炭灰份与熟球产量之间对应的关系, 给出产量预测模型。在已知各原料用量和质量的情况下, 预测产量, 调整短期作业计划。

3.1 数据挖掘技术的选择

与通常的数据库管理系统中的数据(如商场购物数据, 银行储蓄、贷款数据等) 相比, 生产过程中产生的数据有着明显的特点, 主要表现在以下几个方面^[12]:

1) 重复性

生产过程一般是一批一批地生产产品的, 设备运转具有明显的周期性, 若数据采集是按一定的时间间隔进行, 则每周期采集的数据量是一定的。

2) 数据的准确性

生产过程中的数据是直接采集的不落地数据, 因此数据准确性高, 处理过程中不必考虑数据缺省等情况。

3) 噪音数据少

过程控制中考虑到成本问题, 采集的信号一般都是经过精心选取的, 因此对过程控制没有直接影响的信号通常是不会被采集的。

基于以上的数据特点选择统计类数据挖掘技术进行产量预测模型的挖掘。

3.2 统计类数据挖掘技术

数据挖掘是一个交叉学科领域，受多个学科影响，包括数据库系统，统计学，机器学习，可视化和信息科学。此外，依赖于所有数据挖掘方法，以及可以使用的其他学科的技术，如神经网络、模糊和粗略集理论，知识表示、归纳逻辑程序设计或高性能计算。依赖于所挖掘的数据类型或给定的数据挖掘应用，数据挖掘系统也可能集成空间数据分析、信息检索、模式识别、图象分析、信号处理、Web 技术、经济、商业、生物信息学或心理学领域的技术^[1]。

统计是数据搜集和描述数学的一个分支。在统计中总要涉及数据，并且常有足够多的数据使得普通人无法明了所有的数据。对于一般人而言，处理数以万亿比特的数据，且要清楚数据的意义和从数据中归纳出模式，其难度是可想而知的。因此必须借助于数学模型为手段，对这些数据进行归纳、推断和预测，寻找数据间的模式。所谓数学模型，就是根据社会现象的内在、外在因素变量及其相互关系，进行抽象和假设，构造一个或一组反应数量关系的数学方程式。利用数学模型，揭示事物的内部结构，分析变量间的相互关系，进行统计推断和预测。

统计学在数据样本选择、数据预处理及评价抽取知识的步骤中有非常重要的作用。以往许多统计学的工作是针对数据和假设检验的模型进行评^[8]，很明显也包括了评价数据挖掘的结果。

统计作为数据挖掘中的一种技术应用是成功的，就是因为统计技术是对同样类型问题的在同样情况下的应用，例如预测、分类和发现。

统计分析技术中使用的数据挖掘模型有线性分析和非线性分析、回归分析、逻辑回归分析、单变量分析、多变量分析、时间序列分析。最近邻算法和聚类分析等技术。

作为统计类数据挖掘技术还涉及一般数据库中的聚集函数、数据的度量、数据分布的图形、数据的趋势、数据的最近邻和数据的聚类等。聚集函数中的 count, sum, avg, max, min 等，以及对数据进行度量的中心趋势、数据的离散度和统计类的图形表示工具都已成为一些成熟的数据挖掘技术^[2]。

3.3 数据挖掘的方法

数据挖掘需要设计出一个计划，并能够按计划将最初的主意变成最终的现实。一个数据挖掘计划的组成如图 3.1 所示^[2]。

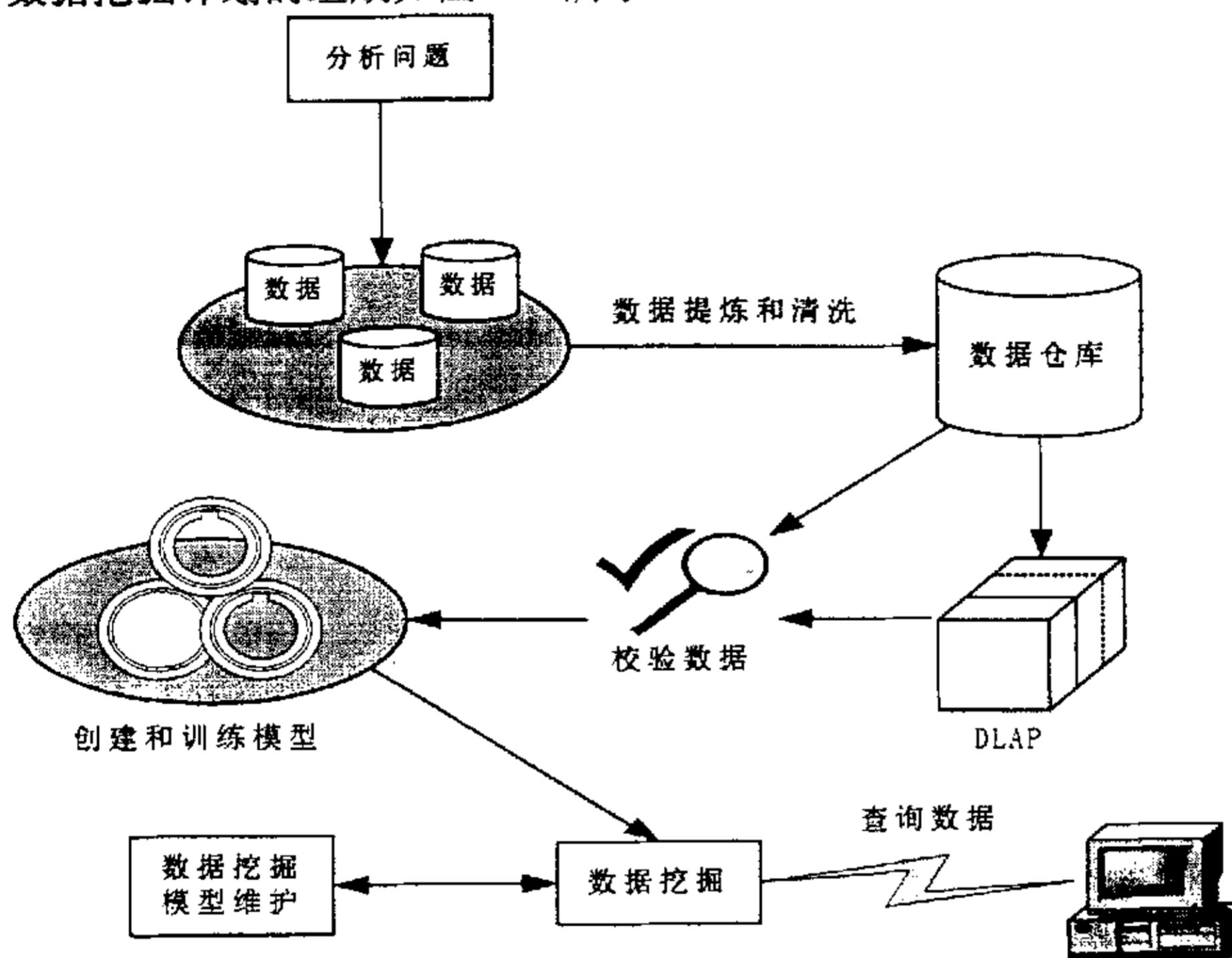


图 3.1 数据挖掘方法

1) 分析问题

我们首先对源数据库进行评估以确认其是否符合数据挖掘的标准。数据的质量和充足是决定数据是否合适的首要因素。然后对数据挖掘的预期结果仔细分析以确认已有数据是否确实能够得出这一类别的信息。

2) 提取和清洗数据

数据最初是从自身本源提取的，比如 OLTP 数据库，文件文本，Access 数据库，以及电子表格。提取后的数据放在一个结构上和数据库模型兼容的数据仓库中。通常，要用数据转换服务器提取数据，以一个统一的格式清洗那些不一致的、不兼容的数据。

3) 校验数据

提取和清理数据后对数据进行检查，以确保所有的数据是都已存在并且完整。

4) 创建和调试模型

当算法应用于模型，即产生了一个结构。浏览产生的数据，与源数据进行比较，确认它对于源数据中事实的准确代表性。

5) 查询数据挖掘模型数据

一旦合适的模型创建并生成了，该数据就可用于决策支持了。该过程通常使用 VB 或 ASP 通过 OLE DB for Data Mining Provider 写成的前端查询，也可以选择使用能够理解的第三方报表工具。

6) 维护数据挖掘模型的有效性。

数据挖掘模型组装好后，随着时间的流逝，初始数据的特征，如粒度或有效性，都可能会发生改变。比如原先的关于原料的质量指标增加或减少了，也许生产工艺改进了，原模型已不再有效。

此外，在进行模型的评估和错误的避免中，数据的质量尤为重要。

因为数据挖掘是通过一组数据来创建现实中的范例，所以模型的好坏在很大程度上依赖于训练数据的质量。如果实例的属性与实际应用中的分布状态很接近甚至相同，那么这个模型对于未来事件的预测是有用的。基于此，建模过程中就要确保用于训练的数据能够精确地反应客观世界中的数据。

1) 数据过密

运用数据挖掘最大的收益之一是可能会由此发现数据中的一些以前未曾注意到的模式。数据挖掘的这种能力使它会发现这样一些模式——这些模式在充斥实例的领域内看来好像比较符合逻辑，但实际上却纯属巧合。

解决数据过密的方法有三种：决策树修剪法（Pruning）、X 平方分析法（Chi-squared Analysis）及交叉验证法（Cross Validation）。

● 决策树修剪法

采用决策树算法对实例属性信息进行评估，排除对预测属性影响较小的一些属性后建模，也可以在发现问题后修改所产生的模型。

● X 平方分析法

X 平方分析法是最常用的一种判断所列实例属性是否恰当的方法。它的公式表达如下： $c^2 = [(期望值 - 实际值)]^2 / 期望值$

- 交叉验证法

交叉验证法通过反复验证已知数据的子集来考察所建模型是否得当。在验证过程中不断重新修正，测试这个模型，直到所得结果与客观情况相符合为止

- 2) 数据过疏

另一个会引起预测误差的原因是数据过疏。当数据挖掘模型没有足够多的属性来支持时，数据过疏现象就有可能发生。尤其是当一些可能对预测结果产生重大影响的属性由于种种原因被忽略掉时，就更容易产生数据过疏。

- 3) 测试数据模型的数据准备

为了确保实际挖掘模型能够真实地反映发展趋向，必须准备一些数据来测试模型的可行性。为此，准备以下三种类型的数据：

- 训练用数据集

在构建数据挖掘地初始模型时，用到训练数据。这个数据集有一些偏差或额外地特性，但是它确实如实反映了用来数据挖掘地样本。

- 测试用数据集

训练用数据集在对模型进行训练地过程中逐渐被校正后，它随之就变成了测试用数据集。也就是说，一旦确信样本中地实例数据集足以能够保证建立一个精确地模型，它们就可以作为输入来预测其他模型，以检验其准确性。测试数据的修正是一个反复的过程，直到测试数据集所建立的模型能令人满意地准确地反映真实的数据为止。

- 评估用数据集

评估用数据集就是从同一个数据种群中抽去的作为评估用的数据集，我们用它的属性值来检验由前两种数据所建立的数据挖掘模型的准确性。

3.4 熟球产量预测模型

3.4.1 属性的初步确定

球团的生产工艺流程如图 3.2 所示。

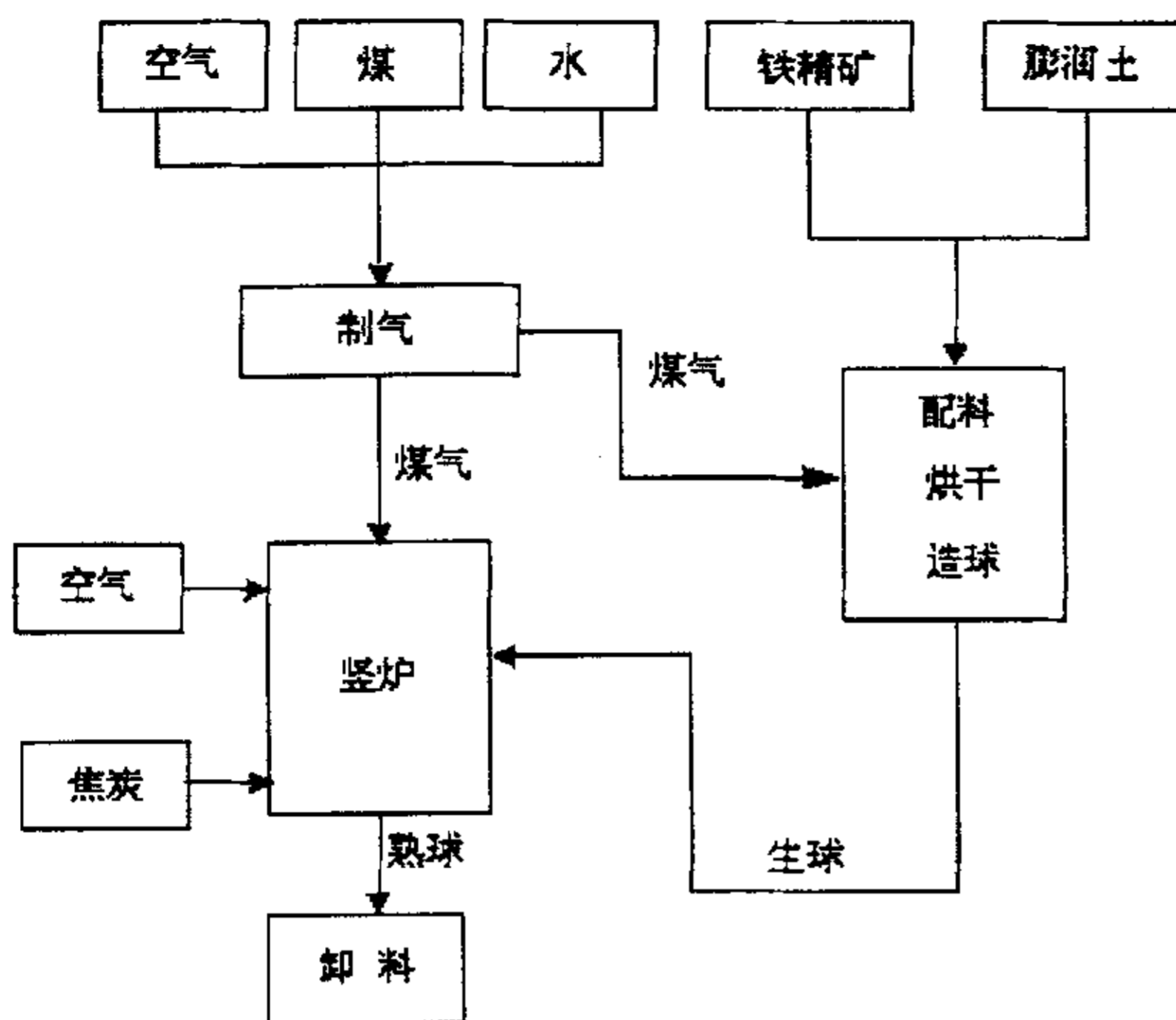


图 3.2 球团生产工艺流程图

由球团厂按生产工艺分为三个工段，它们分别是制气、造球、竖炉工段。

根据该工艺流程可知，膨润土、铁精矿和焦炭为影响熟球产量的主要原料，煤气为燃料，虽然竖炉和烘干工段都消耗大量的煤气，但在这些工艺中主要发生的是物理反应，煤气为燃料，主要用来控制炉温，不影响熟球产量。目前厂方用的膨润土取材方便且价格低廉（相对铁精矿和焦炭的价格要低很多），而且在配料工艺过程中铁精矿和膨润土是工人根据长期操作经验以一定的配比送料的。由于本文的主要目的是在已知各原料用量的情况下，预测产量，调整短期作业计划。在此选取原料铁精矿和焦炭用量为熟球产量预测初始模型中的自变量，熟球产量为因变量。

3.4.2 线性回归模型

一般而言，回归是研究自变量与因变量之间关系的分析方法。其目的在于根据已知自变量来估计和预测因变量的总平均值^[13]。例如，铁精矿用量和焦炭用量与熟球产量有着依存关系，通过对这一依存关系的分析，实现熟球产量的预测或根据熟球计划产量，可以合理安排铁精矿和焦炭的用量，在保证产量的基础上节约原料。

线性回归是最简单的回归形式。双变量回归将一个随机变量 Y （称作响应变量）看作为另一个随机变量 X （称为预测变量）的线性函数，即

$$Y = \alpha + \beta x \quad (3.4.1)$$

其中，假定 Y 的方差为常数， α 和 β 是回归系数，分别表示直线在 Y 轴的截距和直线的斜率。这些系数可用最小二乘法求解。这使得实际数据与该直线的估计之间误差很小。

这种模型的最普遍形式得到的响应变量 y 的预测值 \hat{y} 也是预报变量的 x_j 线性组合：
$$\hat{y} = a_0 + \sum_{i=1}^p a_i x_i \quad (3.4.2)$$

当然实际上我们通常不能完美的预测出响应变量，因此普遍的目标是预测出 y 在预报变量的每个向量位置处所取的均值——所以 \hat{y} 是我们对 y 在 $x = (x_1, \dots, x_p)$ 点的均值的预测性估计。这种形式的模型被称为线性回归模型 (linear regression model)。在最简单的情况中仅有一个预报变量 (单一回归)，这时在响应变量合预报变量所跨越的空间中可以得到一条回归直线 (regression line)。更一般的情况是多重回归 (multiple regression)，这时是一个回归平面 (regression plane)，这种模型是最古老、最重要而且应用最广泛的预测模型形式。之所以如此的一个重要原因是这种模型具有明显的简洁性：简单的加权求和既易于计算又易于理解。另一个非常有说服力的原因是它们在很多情况下都可以达到非常好的性能——即使是对于我们有足够把握认为预报变量合响应变量不是线性关系的情况。不过这并不是空穴来风，而是有道理的：如果我们把连续的数学函数用泰勒级数展开，那么我们经常会发现次数最低的项——线性的项——是最重要的，因此可以使用线性模型得到最好的简单近似。

选取的模型恰好完全正确的情况是非常少的。对于数据挖掘来说更是如此，因为在数据挖掘中模型通常都是实验性的，而不是建立在理论基础上的。此外，我们的模型可能没有包含理想预测所必需的所有预报变量 (很多变量可能还没有被测量出来或者甚至是不可测量的)；或者可能没有包含预报变量的某个函数 (或许不仅需要 x_1 还需要 x_1^2 ，或者需要预报变量相乘，因为它们对 y 的影响是相互作用的)；而且任何情况下测量值都不是十全十美的。所以对变量 y 的预测会存在关联误差，从而使每个向量 (x_1, \dots, x_p) 是和可能 y 值的分布相联系的，就像我们上面所指出的。

所有这些问题意味着样本中实际 y 值会和预测出的值不同。观察到的和预测出的值之间的差异被称为残差 (residual)，我们把它表示为 e ：

$$y(i) = \hat{y}(i) + e(i) = a_0 + \sum_{j=1}^p a_j x_{j(i)} + e(i), \quad 1 \leq i \leq n \quad (3.4.3)$$

按照矩阵表示, 如果我们用向量 y 表示训练样本中的 n 个对象的观察到 y 测量值, 用 $n \times (p+1)$ 的矩阵 X 表示测量 n 个对象得到的 p 种预报变量值 (加入额外的一列 1 是为了和模型中的截距项 a_0 对应), 那么我们可以根据前面的模型把观察到响应值和预报测量值间的关系表示成:

$$y = Xa + e \quad (3.4.4)$$

其中 y 是 $n \times 1$ 的响应值矩阵, $a = (a_0, \dots, a_p)$ 表示 $(p+1) \times 1$ 的参数值向量, $e = (e(1), \dots, e(n))$ 是包含残差的 $n \times 1$ 向量。显然我们需要选取模型中的参数 (向量 a 中的 $(p+1)$ 个值) 使得到的预测尽可能地准确。换个角度来说, 我们必须以某种方式找到对 a_j 的估计使分歧 e 最小化。为了实现这个目的, 我们合并 e 中的元素以得到一种可以最小化的单一数学尺度。人们已经提出了很多种合并 $e(i)$ 的方法, 但是到目前为止最流行的方法是对它们的平方求和——也就是误差平方和评分函数。这样, 我们只要求出使下式最小化的参数向量 a :

$$\sum_{i=1}^n e(i)^2 = \sum_{i=1}^n \left(y(i) - \sum_{j=0}^p a_j x_j(i) \right)^2 \quad (3.4.5)$$

在这个表达式中, $y(i)$ 是在第 i 个训练样本点观察到的 y 值, 并且 $(x_0(i), x_1(i), \dots, x_p(i)) = (1, x_1(i), \dots, x_p(i))$ 是这个点的预报变量向量。出于很明显的理由, 这种方法被称为最小二乘法 (least squares method)。简单起见, 我们把使上式最小化的参数向量表示为 (a_0, \dots, a_p) 。(当然如果我们使用某种符号指出它是一种估计, 比如 $(\hat{a}_0, \dots, \hat{a}_p)$, 那么会更准确, 但是我们的表示更加简洁。) 如果使用矩阵形式, 那么可以证明使公式 3.4.5 最小化的参数值为: $a = (X^T X)^{-1} X^T y$

在线性回归中通常把参数 a 称为回归系数。一旦估计出了这些参数, 就可以把它们代到公式 3.4.2 中进行预测了。对于预报变量的向量 x_k , 可以用 $\hat{y}_k = x_k^T a = a^T x_k$ 预测出它的 y 值 \hat{y}_k [13]。

下面详细阐述运用线性回归作出产量预测的线性模型的过程。

在球团生产调度数据库中保存了有关各原料用量和熟球产量的每日统计数据, 该数据由调度人员每日统计后输入并保存制生产调度的关系数据库。

在关系数据库模式中的一个关系模式是描述一个实体的属性的集合 $R =$

$\{A_1, A_2, \dots, A_n\}$, 每个属性有一个取值范围即域, 关系模式上的一个元组是 $\text{dom}(A_1) \times \text{dom}(A_2) \times \dots \times \text{dom}(A_n)$ 的一个元素, $\text{dom}(A_1), \text{dom}(A_2) \dots \text{dom}(A_n)$ 分别表示属性 A_1, A_2, \dots, A_n 的域, 关系模式上的关系 r 是元组的一个有限集合, 因此可以把属性看成是一个随机变量。一个关系模式是一个随机向量, 一个元组是随机向量的观察值, 一个关系是随机变量集的一个样本^[18]。因此可以利用统计理论来挖掘数据库中的统计关系, 例如, 考虑关系模式: 熟球产量 (铁精矿用量, 焦炭用量)。

一个属性间的统计相关性是指一个属性 Y 依赖于另一个属性集 X , 属性 Y 称为依赖关系属性, 属性集 X 称为解释属性。

假设铁精矿用量与焦炭用量和熟球产量有关并存在近似的线性关系, 则可设依赖属性 Y 为熟球产量。解释属性 X_1, X_2 为铁精矿用量与焦炭用量, 利用线性回归模型:

$$\hat{y} = a_0 + \sum_{j=1}^p a_j x_j \quad (3.4.2)$$

$$\text{即 } Y = a_1 X_1 + a_2 X_2 + a_0, \quad (3.4.6)$$

回归系数 a_0, a_1, a_2 的估计可由大小为 3 (关系地元组数) 的样本获得, 即

$$\hat{Y} = \hat{a}_1 X_1 + \hat{a}_2 X_2 + \hat{a}_0 \quad (3.4.7)$$

模型的显著性检验: 设

$$U = \sum_{i=1}^3 (\hat{Y}_i - \bar{Y})^2, \quad (3.4.8), \quad L_{YY} = \sum_{i=1}^3 (Y_i - \bar{Y})^2 \quad (3.4.9)$$

$$Q = L_{YY} - U, \quad (3.4.10), \quad \text{在模型不显著性的假设下,}$$

$$F = \frac{U/p}{Q/(n-p-1)} \sim F(p, n-p-1), \quad (3.4.11)$$

在给定置信水平 α 的情况下, 设 F 的临界值为 f_α , 若 $F > f_\alpha$, 则原模型事显著的。

如果模型 3.4.6 是显著的, 并不说明每一个属性变量是显著的, 若属性变量 X_i 对 Y 的作用不显著, 则从 3.4.6 中将属性变量 X_i 剔除, 剔除 X_i 后原模型 3.4.6 将重新调整。在属性变量 X_i 对 Y 的作用不显著的假设下,

$$F_i = \frac{\hat{a}_i^2 / c_{ii}}{Q / (n-p-1)} \sim F(1, n-p-1) \quad (3.4.12)$$

其中 c_{ii} 是矩阵 C 的元素, $C = (c_{ij})_{p \times p}$, $C = X^T X$, $X = (X_{ij})_{n \times p}$, 在给定置信水平

α 的情况下, 设 F 的临界值为 f_α , 若 $F_i > f_\alpha$, 则属性变量 X_i 对 Y 的作用是显著的。否则, 属性变量 X_i 对 Y 的作用不显著, 从 3.4.2 中将属性变量 X_i 剔除, 剔除 X_i 后原模型 ## 的系数将重新调整如下, 其中 $j \neq k$

$$\hat{\alpha}_i^* = \hat{\alpha}_i - \frac{c_{ij}}{c_{kk}} \hat{\alpha}_k \quad (3.4.13)$$

重复以上过程, 直到所有的属性变量都不能剔除为止, 这样得到的最终方程称为最优回归方程。

最终得熟球产量预测模型: $Y = 0.764X$

其中 X 为铁精矿用量, Y 为熟球产量, 且 $X \in (490, 1787)$, $Y \in (374, 1365)$, 单位均为吨。将产耗数据表中 584 组铁精矿班用量和熟球班产量数据元组作对应的散点图, 得图 3.3。

熟球班产 (单位: 吨)

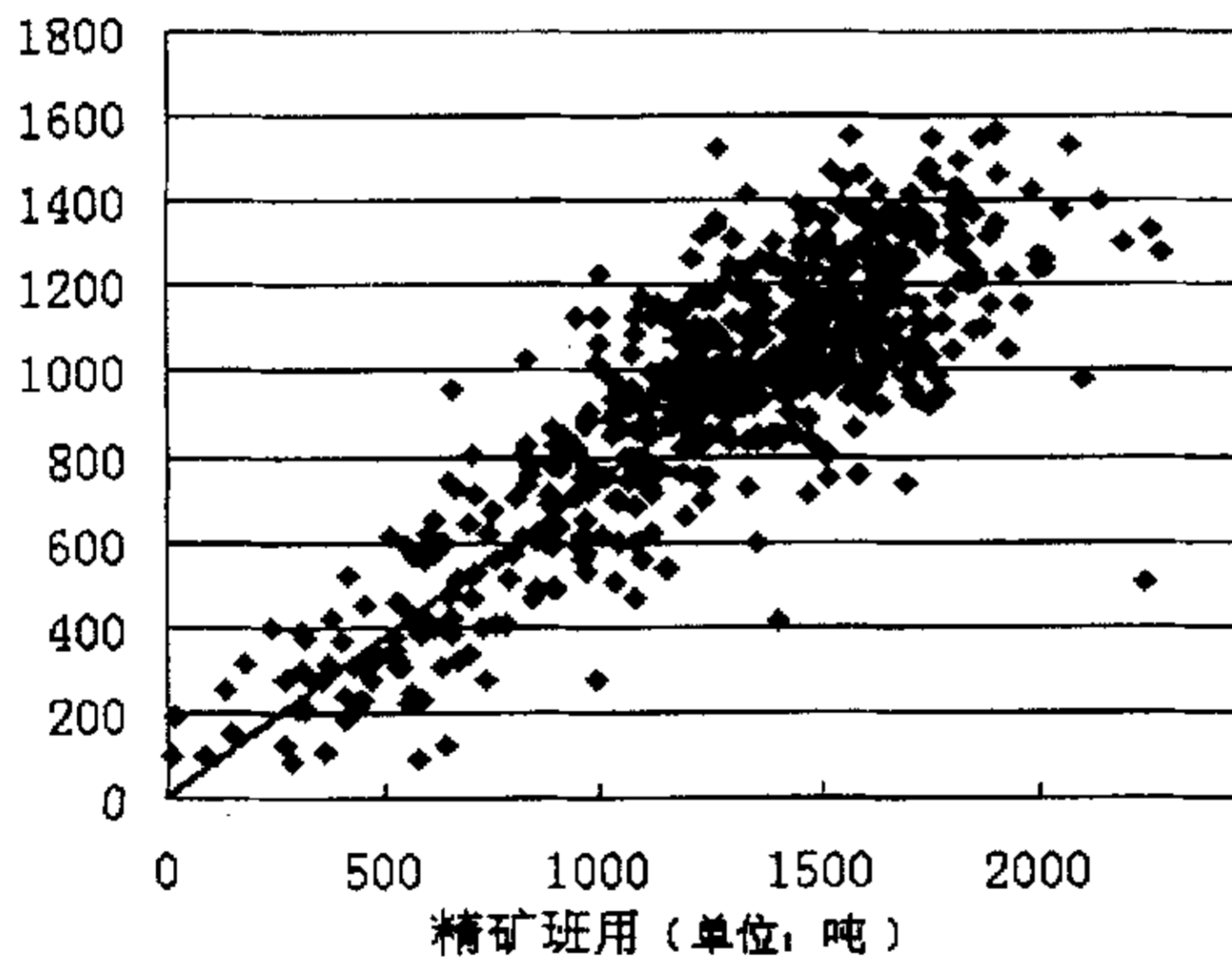


图 3.3 铁精矿班用与熟球班产散点图

由图 3.3 可见, 铁精矿用量与熟球产量基本上成线性关系, 但对于一定的铁精矿用量, 熟球产量仍有较明显的波动。上面的线性统计挖掘可能出现了由数据过疏引起的预测误差。当数据挖掘模型没有足够多的属性来支持时, 数据过疏现象就有可能发生。尤其是当一些能对预测结果产生重大影响的属性由于种种原因被忽略掉时, 就更容易产生数据过疏。这里, 我们忽略了膨润土胶质价、铁精矿品位和硫份等可

能对熟球产量有较大影响的一些原料质量指标。

3.5 原料质量对产量的影响因素

多元回归模型中，用多个自变量说明或预测因变量 y ，再建立这一模型时，可能会遗漏了某些能影响 y 变化的自变量，也可能加入了一些与 y 无关的变量。这些都会对回归拟合产生不利的影响。

- 1) 在回归方程中，若遗漏了应加入的变量，将使所有回归系数估计量产生偏差。
- 2) 若加入了不该加入的变量，将加大所有的回归系数估计量的方差。

所以在多元的回归模型中需要有变量选择的功能。

将该线性模型运用到现有的大量的关系元组中进行验证时，拟合度不好。当铁精矿用量一定时，熟球的产量波动很大。关系模式：熟球产量（铁精矿用量）忽略了可能对熟球产量有较大影响的一些原料质量指标。

球团生产的主要原料是铁精矿、膨润土和焦炭。每种原料的质量指标都有几个。如膨润土的质量指标有：水份、胶质价、膨胀容、吸兰量。如果把每一个质量指标都放入产耗统计模型，那么会导致模型的臃肿和统计的复杂而整体拟合度低。因此，在进行产耗统计之前，先要选择合适属性，初步确定要统计的关系模式。文中，先通过聚集函数对原料质量的各个指标作简单统计分析。

在许多数据库中都包含常用的聚集函数，例如 `count`，`sum`，`avg`，`max`，`min` 等。这些函数在数据挖掘中主要发挥概要统计作用。如果数据对象的值和某个权重有关，即值的大小需要考虑值的意义、重要性或频率，就不能简单的用算术平均来度量数据对象的中心趋势，而需要采用加权算术平均值。

在一个简单的样本数据库中，对于布尔型属性可以通过计算数据库中不同布尔值发生的次数，就可创建关于这个属性的柱状图。

这里的对象属性都是数值型的，首先使用聚集函数（`count`，`sum`，`avg`，`max`，`min`）进行统计前数据的清洗，剔除奇异值。

然后我们采用了 SQL Server 的 Analysis Manger 中的决策树方法，对铁精矿用量一定的情况下，以铁精矿硫份，铁精矿粒度，焦炭灰份，焦炭硫份和膨润土胶质

价作为决策属性，以熟球产量作为决策目标得决策树如图 3.4 所示。

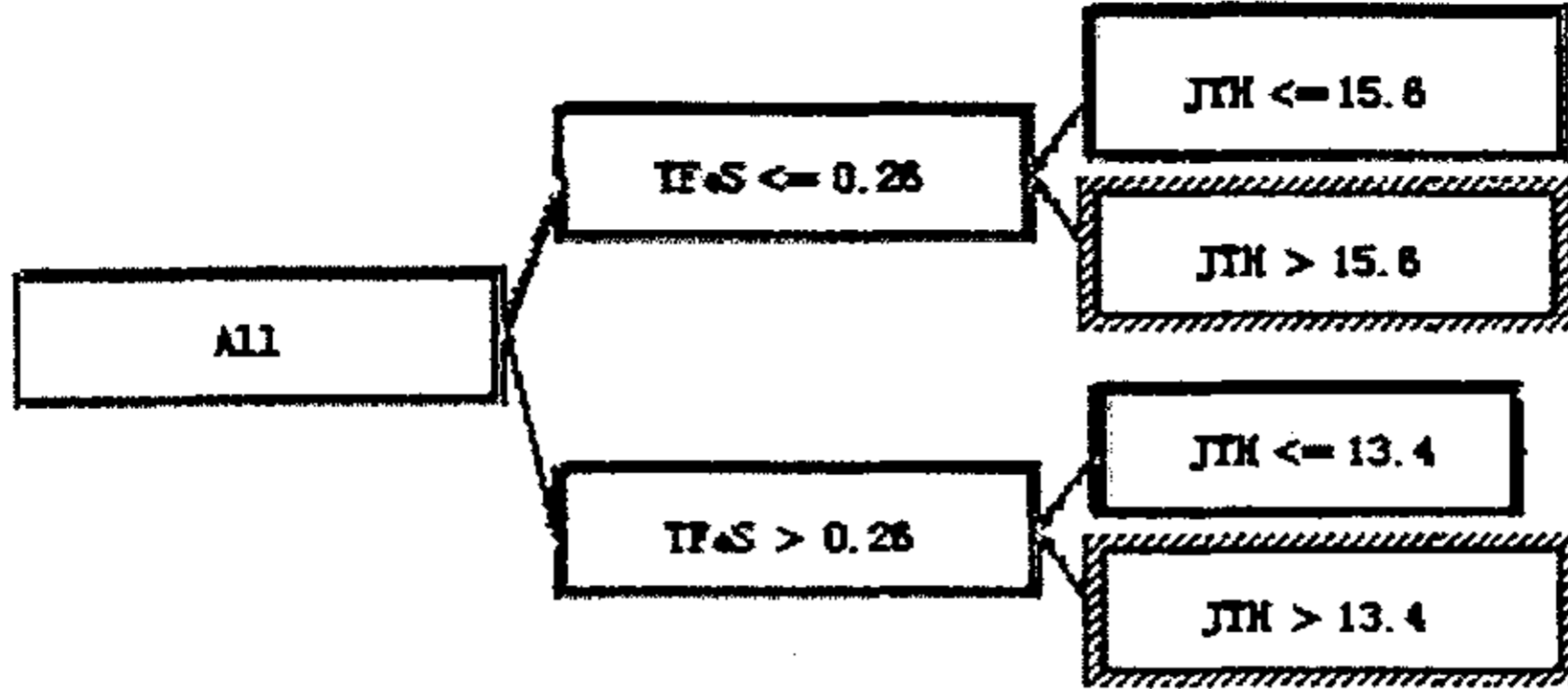


图 3.4 原料质量对产量影响挖掘决策树

由决策树可知，在铁精矿用量一定的情况下，在原料质量指标中，铁精矿硫份和焦炭灰份对熟球产量的影响相对较大。因此建立完善上一节的产量预测模式为熟球产量与铁精矿硫份和焦炭灰份、铁精矿用量之间对应的关系。

而铁精矿硫份和焦炭灰份与熟球产量之间的关系又是如何的呢？首先，我们还是想到了线性关系，可是由铁精矿硫份与产量、焦炭灰份与产量的对应散点图，排除了它们之间是简单线性关系的假设。

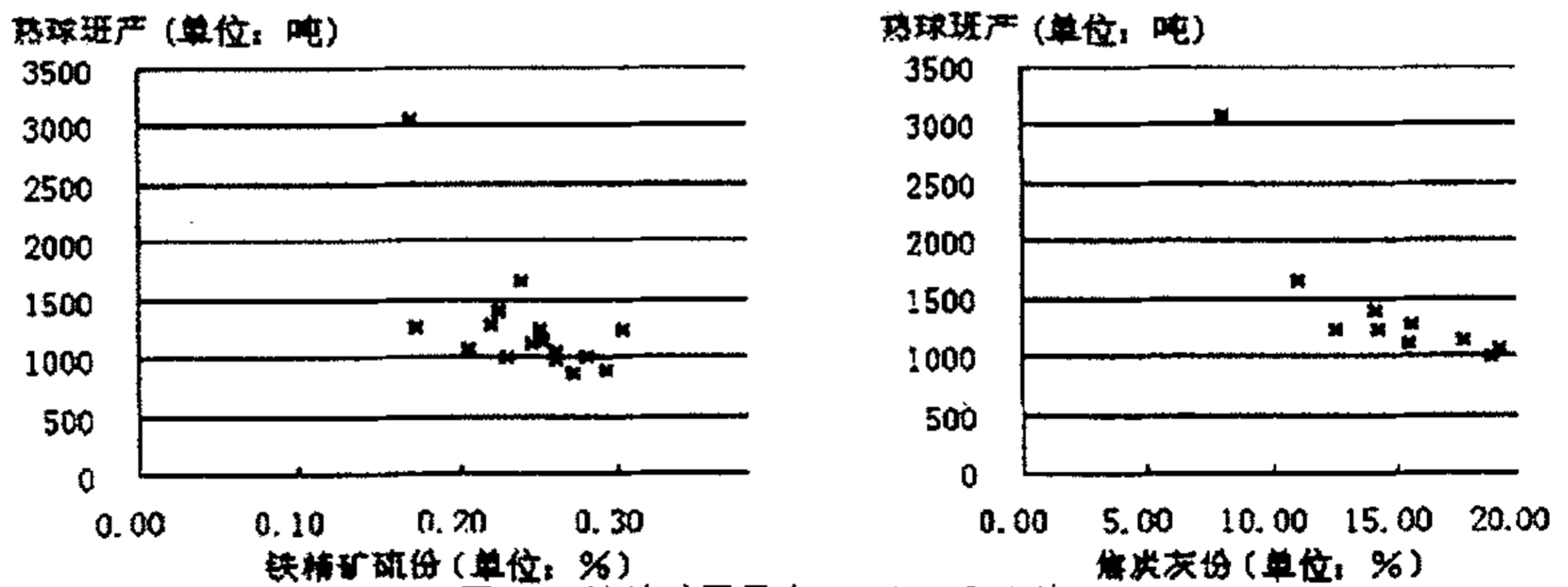


图 3.5 铁精矿用量在 1233-1211 吨

范围内的产量与原料质量指标对应散点图

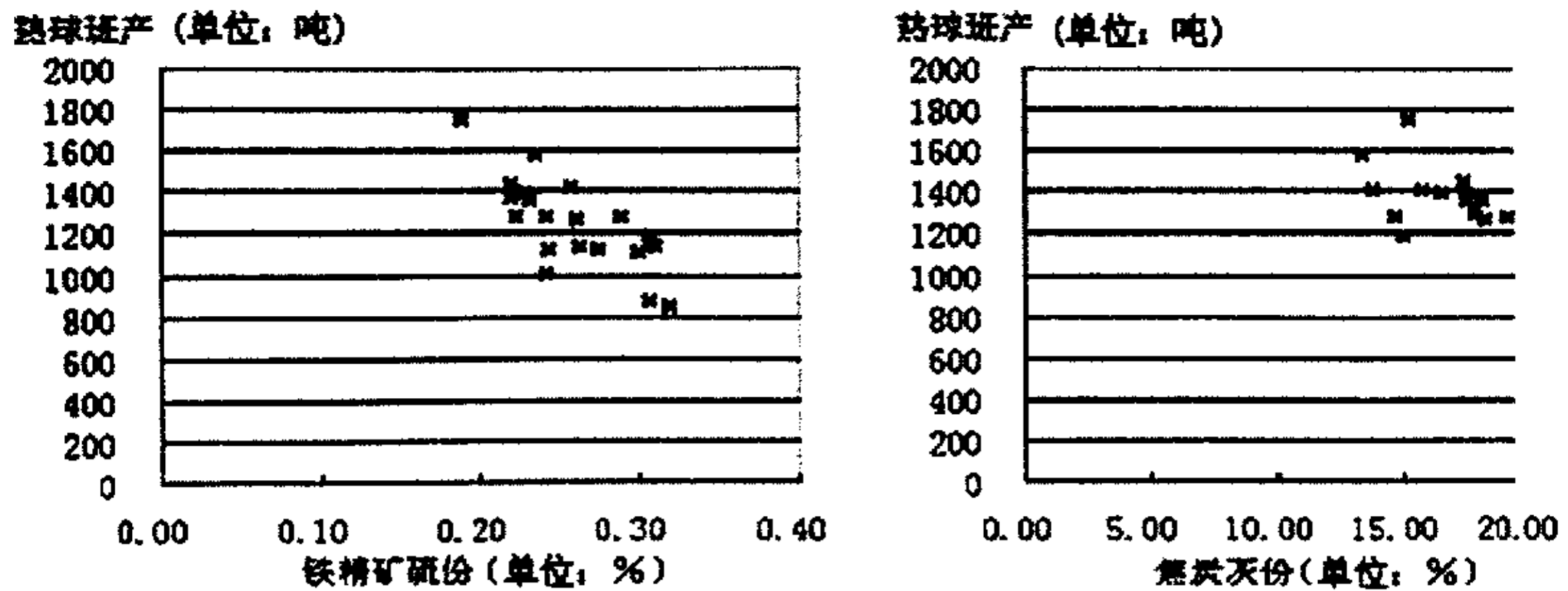


图 3.6 铁精矿用量在 1625—1609 吨

范围内的产量与原料质量指标对应散点图

当判定变量间的关系大致是一条直线时，可以拟合一条直线反应其变动关系。然而在很多情况下，变量间的关系呈曲线形式，即非线性的，这时就应拟合一条曲线来反映变量间的关系。例如，给定的响应变量和预测变量间的关系可用多项式函数表示。通过对变量进行变换，可将非线性模型转换成线性的，然后用最小二乘法求解。非线性回归主要有以下 7 种模型^[2]。

- 1) $y_i = \beta_1 + \beta_2 \frac{1}{x_i} + \varepsilon_i$
- 2) $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i$
- 3) $y_i = \beta_1 + \beta_2 \ln x_i + \varepsilon_i$
- 4) $y_i = \beta_1 + \beta_2 \sin x_i + \varepsilon_i$
- 5) $\begin{cases} y_i = \alpha \beta x_i + \varepsilon_i \\ \ln y_i = \beta_0 - \beta_1 - \beta_2 x_i - \varepsilon_i \end{cases}$
- 6) $y_i = \alpha x_i^b + \varepsilon_i$
- 7) $y_i = \alpha + \beta x_i^2 + \varepsilon_i$

由散点图及决策树，很难看出铁精矿硫份和焦炭灰份与熟球产量之间的是哪种非线性关系，它们之间的关系也许并非以上的直观的非线性模型所能描述。

于是，我们考虑使用神经网络。神经网络属于高度参数化的统计模型，它可以克服一般统计模型的缺点精确地模拟出函数中非常小的不规则性，但同时也要避免过度拟合。下一章，将详细阐述神经网络在优化模型中的应用。

3.6 本章小结

本章在介绍统计类数据挖掘技术和挖掘的方法的基础上，运用线性回归模型和最小二乘法等回归统计挖掘方法和显著性验证模型得出了关于熟球产量与铁精矿用量的线性回归模型。在分析了所得线性模型的缺陷后，加入影响产量的原料质量属性，通过决策树得出在铁精矿用量一定的情况下对熟球产量影响相对较大的原料质量属性为铁精矿硫份和焦炭灰份，建立相对完善的产量预测模型即：熟球产量与铁精矿用量、铁精矿硫份和焦炭灰份之间的对应关系。并由散点图分析得知，它们的关系是复杂的非线性的，由此引出了神经网络的应用。下一章将对神经网络在模型挖掘中的应用做详细阐述。

4 生产计划模型

本章首先考虑到单纯的产量并不体现实际的生产状况，于是提出创建产量与质量相结合的新属性—综合产量，为此对球团厂一年的熟球综合质量进行统计评定，评定的结果为熟球质量较稳定，避免了创建新属性综合产量。随后采用上一章提出的完善后的产量预测模型，并利用神经网络训练和实现产量预测模型，以铁精矿用量、铁精矿硫份、焦炭灰份为网络输入，以产量为网络输出。最后在产量预测模型的基础上，改变模糊神经网络的输入和输出，以产量、铁精矿硫份、焦炭灰份为网络输入，以铁精矿用量、焦炭用量为网络输出，得出生产计划模型。

神经网络属于高度参数化的统计模型，它可以克服一般统计模型的缺点精确地模拟出函数中非常小的不规则性，但同时也要避免过度拟合。在最终的模型输出中，将模糊处理功能引进神经网络，增强了神经网络的输出表达能力，使模型变得更加稳定可靠^[2]。

本章的主要目的是挖掘产量预测和生产计划模型，最终实现产量的准确预测和原料的合理分配，进行生产调度。

4.1 综合产量与质量

上一章最终给出的产量预测挖掘模型为熟球产量与铁精矿用量、铁精矿硫份和焦炭灰份之间对应的关系。该模型在一旦实现，可以在已知各原料用量和质量的情况下，预测产量，调整短期作业计划。

而单单的熟球产量并不体现实际的生产状况，如果熟球质量较低尤其是等外品较多，则产量高并不等于生产状况好。因此，将熟球产量和质量相结合的属性才能更贴切的反映生产状况。这里提出综合产量的概念。在所有生产出的熟球中，分一等、二等、三等和等外这四个等级质量，这四个等级的熟球数量在总产量中的比例体现了生产的质量水平。我们给这四个等级赋不同的权值 p_1 、 p_2 、 p_3 和 p_4 ，设四个等级的质量的熟球产量分别 y_1 、 y_2 、 y_3 、 y_4 ，综合产量就是各个等级的熟球产量

华中科技大学硕士学位论文

与对应等级权值的乘积的总和。各等级权值的设定比较复杂，我们先从分析球团厂较长一段时间的熟球综合质量入手。

在球团生产调度数据库中保存了有关熟球质量的每日统计数据，该数据由质检科上报调度人员每日输入并保存。其中包括数据元组形式如表 4.1，根据生产经验，选取抗压强度作为质量衡量的指标。

时间	班次	抗压强度			
		一级	二级	三级	等外
1. 1	丙	57	10	13	20
	丁	53	10	14	23
1. 2	乙	57	13	10	20
	甲	53	10	13	24
1. 3	丁	53	13	11	23
	丙	57	10	13	20
1. 4	甲	50	10	17	23
	乙	50	17	10	23

表 4.1 熟球抗压强度质量表

将每班的抗压检验结果标准化，即分别用各级熟球在总体检验样本中的比例代替其具体的个数。得表 4.2 形式如下：（由于数据太多，文中只列出部分旨在明确其形式）

时间	班次	抗压强度			
		一级	二级	三级	等外
1. 1	丙	0.57	0.10	0.13	0.20
	丁	0.53	0.10	0.14	0.23
1. 2	乙	0.57	0.13	0.10	0.20
	甲	0.53	0.10	0.13	0.24
1. 3	丁	0.53	0.13	0.11	0.23
	丙	0.57	0.10	0.13	0.20
1. 4	甲	0.50	0.10	0.17	0.23
	乙	0.50	0.17	0.10	0.23

表 4.2 熟球抗压强度标准化表

比较以上两表，发现下表数据是上表中对应数据的百分之一，这是由于检验的总个数为定值（也许就是 100），然而这只是大多数情况。有时检验的总个数改变了，两表之间的比例就不一定了。所以以上的标准化是必要的。

熟球按质量分成了一级、二级、三级、等外这四个等级，标准化后得出四个等级的熟球分别在总熟球量中的比例，对改表进行统计分析，得均值为：0.544:0.137:0.124:0.195，波动最大的是等外品的比例，在 0.19 到 0.24 之间成置信度 95% 的分布，可见一级、二级、三级、等外的比例基本上在小范围能波动，球团厂的熟球质量是较稳定的。

在质量稳定的前提下，单独的产量可以代替综合产量来体现生产状况，因此避免了通过产量与质量综合以求得综合产量的复杂步骤。

4.2 基于神经网络的预测模型

由于球团厂的熟球质量是稳定的，所以单纯的产量可以体现并代表综合产量。同时为了模型的简化和实用性，我们的综合产量预测模型还是以单纯的产量为预测目标。

上一章的挖掘结果给出了产量预测挖掘模型即熟球产量与铁精矿硫份、焦炭灰份和铁精矿用量之间对应关系，而且它们之间是复杂的非线性关系，因此我们考虑使用神经网络技术。

神经网络的挖掘过程基本上是将数据聚类，然后分类计算权值。与一般的传统数据处理方法相比，神经网络技术在以下几方面有其独特的优越性，因而很适合非线性和含噪声数据。

- 1) 模糊的数据；
- 2) 需要决定的模式特征不明确；
- 3) 数据本身非线性；
- 4) 随机数据或数据中含有较多的噪声。

神经网络是大量的简单神经元按一定规则连接构成的网络系统。网络能够模拟人类大脑的结构和功能，采用某种学习算法从训练样本中学习，并将获取的知识存储在网络各单元之间的连接权中。

神经网络的学习有两种不同的学习或训练方法，即有指导的训练和没有指导的训练。指导下的学习和训练需要“教师”来进行指导，教师即是训练数据本身，不但包括有输入数据，还包括有在一定输入条件下的输出。没有指导的学习过程指只

有输入而没有输出，网络必须根据一定的判断标准自行调整权重。本文中所采用的是有指导的训练。在有指导的学习方式中，网络将应有的输出与实际输出数据进行比较。网络经过一些训练数据组的计算后，最初随机设置的权重经过网络的调整，使得输出更接近于实际的输出结果。所以学习过程的目的在于减小网络应有的输出与实际输出之间的误差，这是靠不断调整权重来实现的。网络经过训练后，若认为网络的输出与应有的输出间的误差达到了允许范围，权重就不再改动了，这时的网络可用新的数据去检验。

基于神经网络的数据挖掘由以下两个阶段组成：网络构造和训练，规则抽取。

4.2.1 数据样本的产生

产生训练样本是成功开发神经元网络的第一步，也是十分重要和关键的一部。这包括原始数据的搜集、数据分析、变量选择以及数据预处理。

样本确定原则：

- 1) 要有一个全面的样本空间：由于神经网络存在一个外推的问题，因此基于神经元网络的软测量技术要得到很好的应用应该有一个全面的样本空间为基础，在实际应用中体现为：样本空间中某变量的范围应该达到或接近神经元输入变量中该变量所处的正常范围
- 2) 每个样本在一定范围内应具有代表性
- 3) 样本数量的确定，不宜过多，但也不能太少。太少，没有代表性，太多，则矛盾数据也会越多，在神经网络建模中将导致神经网络无法辨识。

上一章的挖掘结果给出了挖掘的对象即熟球产量与铁精矿硫份、焦炭灰份和铁精矿用量之间对应的关系模型。而通过上一节的分析，球团厂的熟球质量基本稳定，单纯的产量基本上体现了生产状况，通过对产量的预测，可以指导调整短期生产计划，并为球团生产的长期计划的制定提供依据。

确定网络输入（铁精矿用量、铁精矿硫份、焦炭灰份）和输出（产量）变量后，将这些变量对应的数据汇总到一个数据表中，使用统计聚类分析的方法剔除数据中的野点，然后检验数据是否存在周期性和固定变换趋势。经统计检验，该数据表中数据不存在周期性和较明显的固定变换趋势。因此，可以随机的抽取其中的三分之二用于网络训练，其余用于测试。

产生数据样本的最后一步是尺度变换,尺度变换常常将它们变换到 $[-1, 1]$ 或 $[0, 1]$ 的范围。

4.2.2 网络构造和训练

这一阶段根据数据集中的属性数目和类数目以及数据特性,选择合适的编码方法,构造一个多层的前向神经网络(一般采用三层BP网络即可)。然后,通过教师示教的方式训练构造好的神经网络。

但是,即使对于只有一个输出单元的网络,随着输入单元的增多,网络各层单元之间的连接数据成倍增长,造成相应的提取规则也将成指数增长,给规则提取造成很大困难,为此有必要对网络进行裁减,网络裁减的目的是在不增加网络的分类错误率的前提下,删除多余的连接和隐层单元,构造一个连接和单元数自相对较小的网络,以利于抽取简明、可理解的规则。裁减算法可采用递减的探测算法。递减式的探测算法是:从一个很大的网络结构开始,在训练过程中,根据特定的问题的需要,逐渐减少结构的各个部分,直到找到能解决问题的网络结构为止。但该算法存在着不一定能找到最优结构的缺点。而遗传算法作为一种全局最优搜索算法对目标函数既不要求连续也不要求该问题可微,仅要求该问题可计算,而且它的搜索始终遍及整个解空间,因此非常容易得到全局最优解。遗传算法的这种特性使得它在进化神经网络结构时,可使进化后得到的神经网络具有结构简单、性能优良的优点。遗传算法已越来越多的应用到神经网络的进化中,以下提出一种用遗传算法进化神经网络的思路方法:

- 1) 随机产生多个编码串,每个编码串对应了一种网络结构,这些编码串构成一个群体;
- 2) 将上述群体中的每个个体解码得到对应的神经网络结构,用数据集中的训练样本分别对这些结构的神经网络进行训练;
- 3) 根据训练的结果或其它策略确定每个个体(即每一种网络结构)的适应度值(它表明了该网络结构性能的好坏);
- 4) 选择若干适应度最大的个体,直接遗传给下一代;
- 5) 利用交叉和变异等操作算子对当前一代进行处理,产生下一代群体;
- 6) 重复步骤(2)~(5),直到当前一代群体中的某个个体(对目一个结构)

满足要求为止，此时找到了一个结构最简单、性能最优的神经网络模型。

由于追求的目标是网络结构的最简化，因此，适应度值定义为：在网络的学习错误率一定的条件下适应值为网络连接数得函数。

4.2.3 规则抽取

这一阶段从进化后的相对简单的网络中抽取分类规则或是关联规则，这是较为关键的一步，即如何理解神经网络所学到的知识。目前有两种主要的规则抽取方法：基于搜索的方法和基于学习的方法。前者主要集中于对网络中单个节点所表示的概念的解释，从每个节点中抽取的规则是由与此节点相连的输入节点来表达得，每个节点只表示唯一的概念。基于学习的方法则将抽取规则的过程看作是一个学习过程，其目标概念就是网络计算的输出，其输入就是网络的输入特征，此方法的目的在于抽取反映输入—输出之间映射关系的规则。此方法的优点在于抽取规则的可信度高，同时不考虑网络内部结构，因而应用范围广泛，一般采用该方法。另外，基于神经网络的决策树构造也是一种重要的规则抽取方法。

4.2.4 BP 神经网络训练和验证

本文利用 BP 神经网络的学习能力来实现综合产耗关系模式。以产量、铁精矿硫份、焦炭灰份为网络输入，以铁精矿用量、焦炭用量为网络输出，进行数据训练，从而完善上一章得出的产耗统计模型。

标准 BP 算法由于学习率选择不当，收敛速度很慢慢。训练时间特别长且得不到结果的表达^[27]。本课题对神经元网络的训练和验证都是在 Matlab 中完成的。Matlab 中的 trainbpx 训练函数就是采用了动量法和学习率自适应调整的策略，动量法降低了网络对于误差曲面局部细节的敏感性，有效的抑制网络陷于局部极小。自适应调整学习率有利于缩短学习时间，提供了学习速度并增加了算法的可靠性。

将神经网络输出的值转换成对应预测产量值，再与实际产量比较，得 575 组熟球产量神经网络输出与实际值的对比散点图 4.1，其中三角形点为神经网络输出值对应产量，菱形点为实际产量对应值。

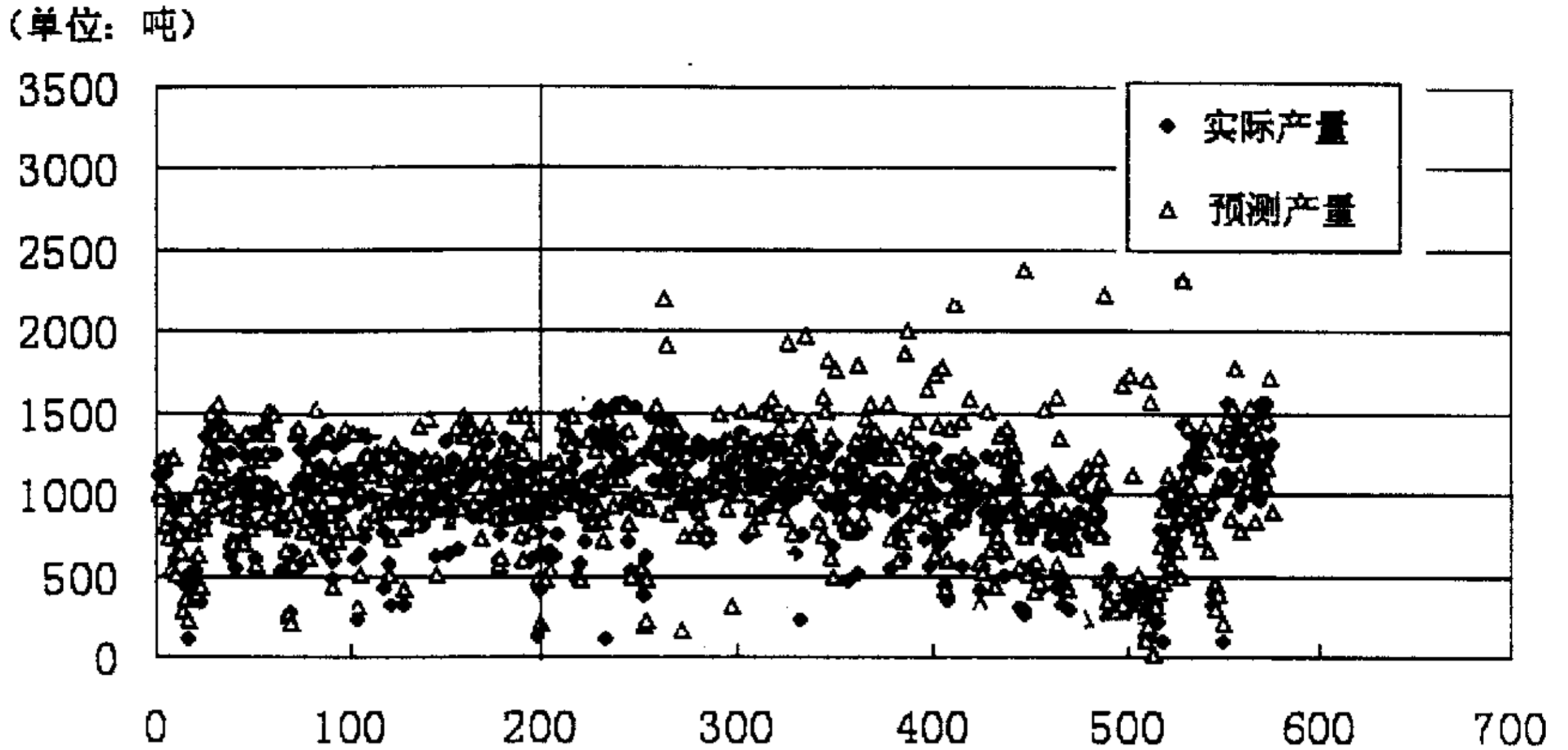


图 4.1 575 组熟球产量神经网络输出与实际值的对比散点图

图中前 200 个星点为训练数据集模拟结果, 后 200 多个点是整个数据样本中除了训练集以外的, 由图可见, 模型对非训练集的数据点的拟合相对误差较大, 其原因主要是由于神经网络的过度拟合和数据的不完全一致性, 由此引入模糊理论来优化原神经网络。

4.3 基于模糊神经网络的预测模型

模糊逻辑与神经网络相结合, 组成模糊神经网络 (Fuzzy Neural Network) 系统。在此先研究如何将模糊理论与神经网络结合起来, 形成模糊神经网络的计算结构 [28]。

最简单也是最有用的方法将模糊逻辑与神经网络结合起来是采用“模糊器” (Fuzzifier)。模糊器是一个预处理器或后处理器, 它可以是一个分开的程序, 也可以是一个处理单元层。模糊器可以接受输入数据, 如直接测量值或部分处理过的数据, 将这些数据转换为模糊逻辑的范畴, 然后再作为普通神经网络的输入。同样, 模糊器也可以接受网络的输出, 并转换为模糊逻辑的形式, 这种情况常用于将网络输出作为专家系统或模糊专家系统的输入。图 4.1 为模糊器与标准的三层后传播网络的结合 [28,29]。

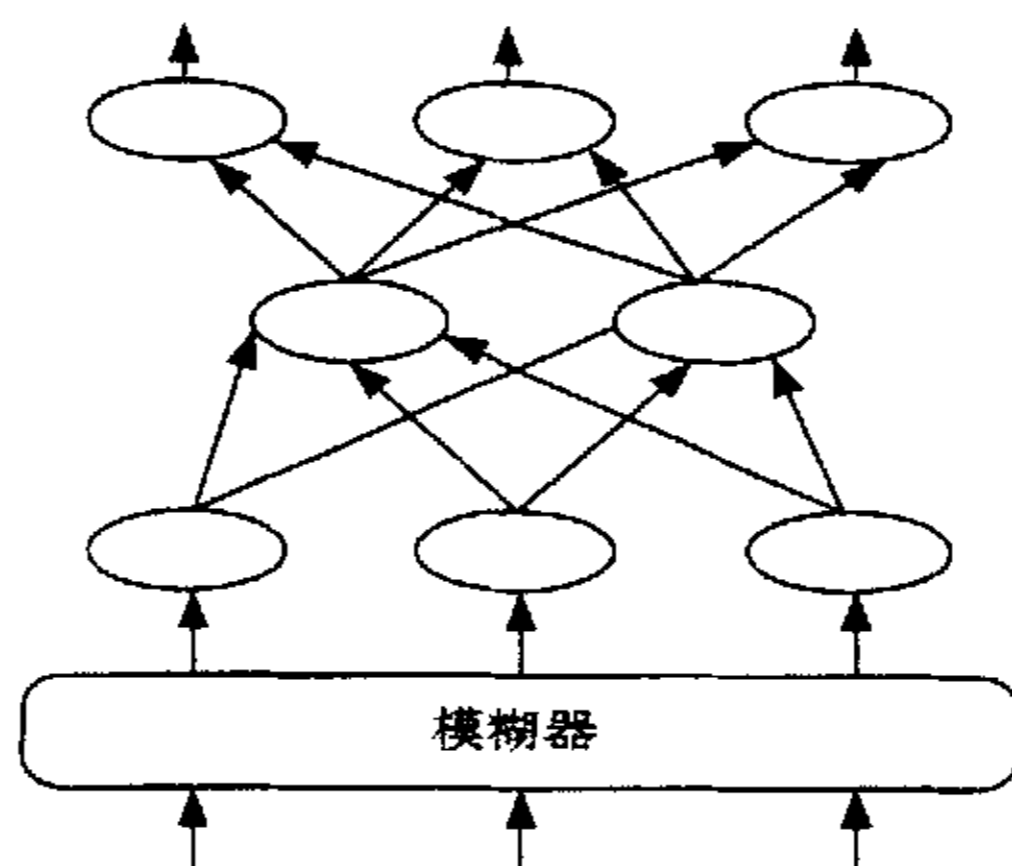


图 4.2 最简单的模糊神经网络结构

模糊神经网络可以处理有关符号推理的不精确问题。如一个变量“产量”可以取一些模糊的分类，如“低”、“中”、“高”、“很高”等等。每个分类都与一定的隶属度函数（Membership function）相关。隶属度函数可以是三角形、钟形或其他一些形式。隶属度可以解释为可能性程度。图 4.2 为与变量“产量”相关的隶属度函数的例子。给定了产量数值，可以从隶属度函数中得到在不同分类中的隶属度。同样，若给定在每一分类中的隶属度，也可以从图中决定合适的产量值。

一般说来，模糊网络可由五层组成，五层包括：输入层、输出层、模糊化层、模糊推理层、去模糊化层。输入层节点数为决策或分类的条件变量数；输出层节点数为决策或分类的目标变量数；模糊化层节点数为条件变量的模糊度划分数的总和；模糊推理层节点数为各条件变量的模糊度划分组合的总和；去模糊化层节点数为各目标变量的模糊度划分数的总和。

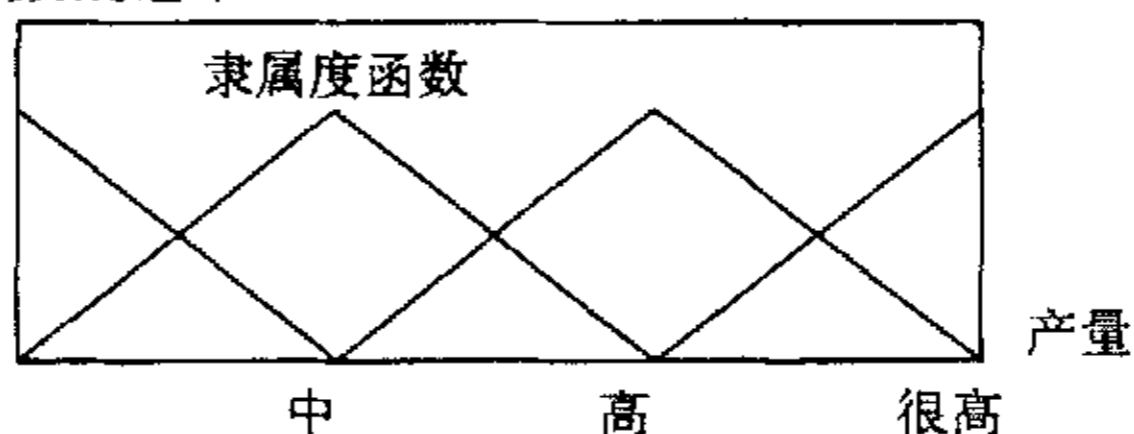


图 4.3 隶属函数图

各步操作中，模糊化是将输入变量的精确数值根据其模糊度划分和隶属度函数转换为模糊度描述。“三角波”隶属度用来定义变量的模糊隶属度。模糊推理运算采用的是 MIN-MAX 推理，其主要步骤是：（1）分别求出条件变量得隶属度（即模糊化）；（2）当有多个条件变量时，AND 关系取条件变量隶属度最小值作为前件部的

隶属度；(3) 前件部的隶属度与后件部的隶属度进行 MAX 运算，得到各推理的结论；(4) 对所有推理结论取 MAX 运算，得到模糊推理的结果。反模糊化是将经模糊推理得到的目标变量的模糊度转换为精确数值，最常用的方法是面积重心法，其计算式为：
$$Z_i = \frac{\sum \mu(Z_i) \times Z_i}{\sum \mu(Z_i)}$$
（即取分步值函数质心处的值为确定性输出）(4.3.1)

以下是两个条件变量和一个目标变量组成的五层模糊神经网络，每个变量划分为 3 个模糊度（如图 4.2 所示）。输入层将输入值传递给模糊层中的模糊单元，模糊层相当于上面说过的模糊器。模糊层将输入值转换为一定的模糊度。中间层的功能与一般神经网络相同，经处理的数据再传送给输出的去模糊层，这一层单元按照模糊度函数将数据再转换为与网络输入值相应的量。从此图中，可得到的模糊规则如下所示：

- (1) 如果 X 为 X1 并且 Y 为 Y2，则 Z 为 Z1
- (2) 如果 X 为 X2 并且 Y 为 Y3，则 Z 为 Z2

从图 4.3 中结构来看，当已知条件变量和目标变量的训练集，并且确定了模糊隶属函数和去模糊方法时，模糊映射关系实际上是学习五层模糊神经网络中模糊推理层与去模糊层之间的权值。模糊神经网络的学习方法与相应的神经网络相同，可以根据条件变量目标变量的数值训练集和误差反向传播（BP）算法进行权重学习。

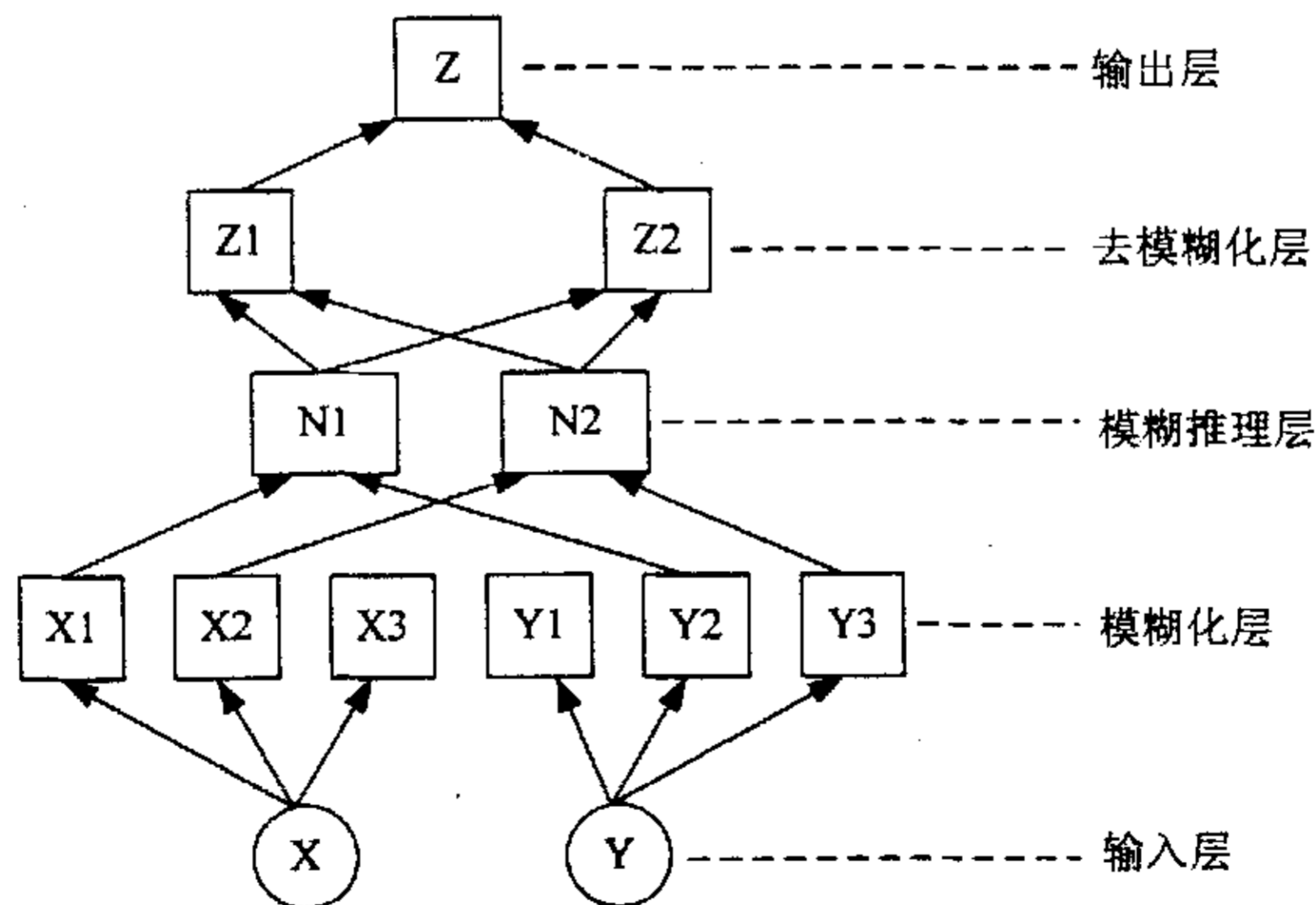


图 4.4 简单的模糊神经网络

- 1) 输入层与模糊化层间的权值为 1，模糊化层节点的输出是输入层变量的各模糊划分的隶属度 $[0, 1]$;
- 2) 模糊化层间与模糊推理层间的权值为 1，模糊推理层节点的输出为模糊推理前件的 AND 操作： $Z = \min(x_k)$;
- 3) 模糊推理层与反模糊层之间的权值为 ω_{ij} 。去模糊层节点的输出为模糊推理后件的 OR 操作： $Z = \max(x_k \times \omega_{ij})$;
- 4) 反模糊层与模糊输出层之间的权应为 1，模糊输出层节点的输出为反模糊化的结果，即：通过面积重心法将模糊推理得到的模糊目标变量的各模糊划分的隶属度转换为数值输出。

在训练学习模糊推理层与反模糊层之间的权值为 ω_{ij} 过程中，可使用 BP 学习算

法使模糊映射得最小均方差变得最小，其中最小均方差定义为 $E = \frac{1}{2} \sum_{i=1}^N (T_i - Z_i)^2$ ，式

中 N 为模糊输出层的节点数； T_i 为模糊目标变量的期望输出； Z_i 为模糊目标变量的实际输出。当模糊推理层与反模糊层之间的权值 ω_{ij} 训练完成后，由于模糊神经网络的节点和权值可用隶属函数和模糊规则来解释，因而从模糊神经网络的权值和节点参数就可直接写出模糊规则和隶属函数（权值 ω_{ij} 可看作是关联强度）。

设想将质量指标数值按一定的分类规则进行属性的模糊化，将其数值转化为枚举值（高、中、低）。对属性铁精矿硫份，焦炭灰份进行模糊化，每个属性都模糊化为高、中、低，采用的隶属函数为 sigmoid 函数，可以构造出一个与图 4.3 类似的模糊数据网络模型。同时也具有一定的鲁棒性和抗噪性，适于挖掘生产数据。

将部分属性转换为模糊逻辑的范畴后再作为普通神经网络的输入进行数据训练，以每日产量和各原料质量的模糊化枚举值为输入，输出各原料用量。利用 Matlab 进行数据训练后得出的模型对验证集的验证结果误差比单纯的神经网络减小了，图 4.5 为模糊神经网络预测模型输出与实际产量的对比散点图，其中星号点为模糊神经网络输出值对应产量值，菱形点为实际产量对应值。

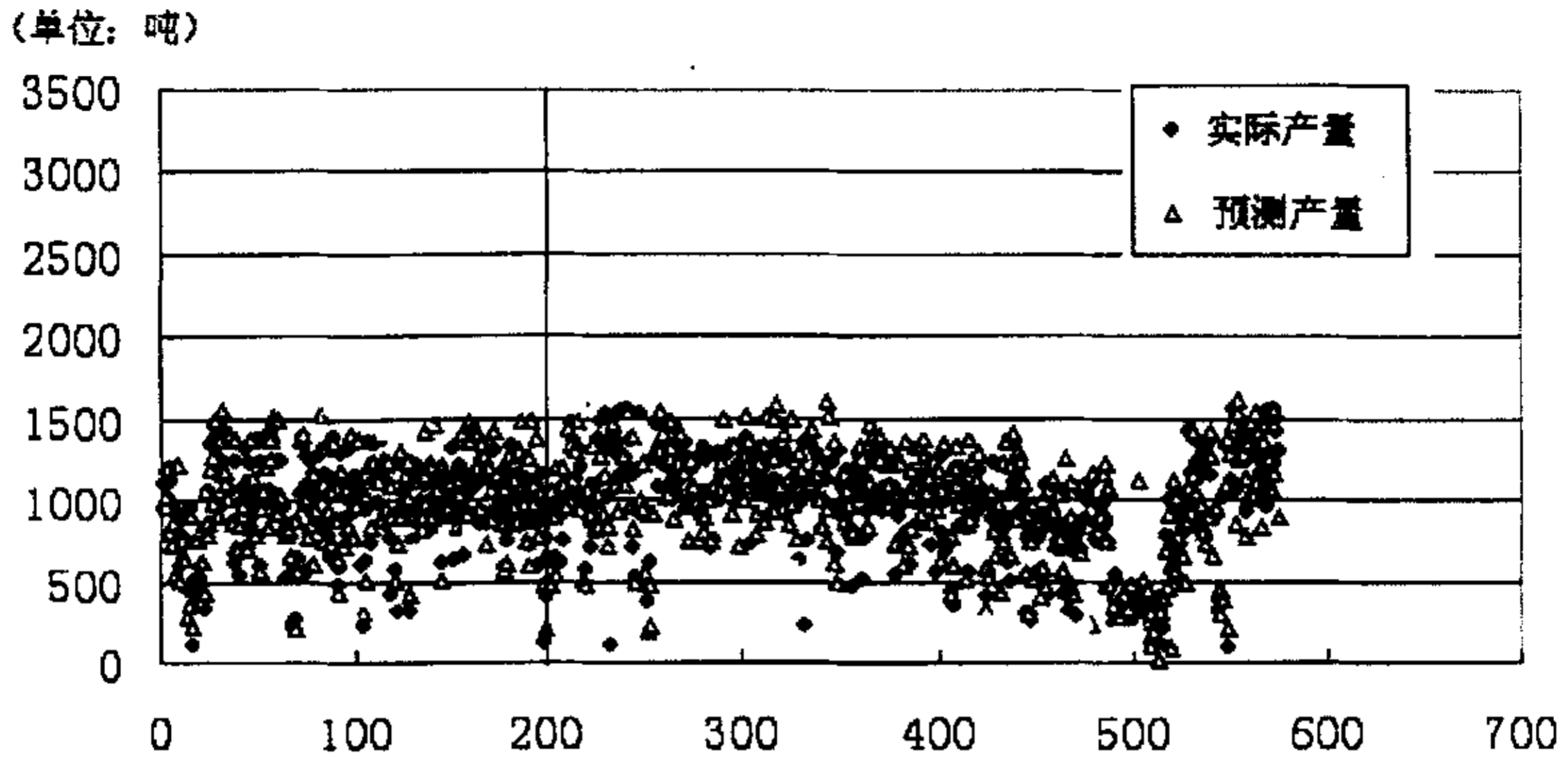


图 4.5 575 组预测产量模糊神经网络输出与实际值的对比散点图

将图 4.5 与图 4.4 比较可见，模糊神经网络的模型对整个样本空间的拟合输出效果比单纯的神经网络好，模型稳定性较好，将该生成模型用于调度系统，可以准确的预测产量，指导短期计划的调整，并为生产计划的制定提供依据。

4.4 生产计划模型

在生产调度中，已知原料用量和质量来预测产量，可以指导短期生产计划的调整，从而为实现长期生产计划合理调度生产。如何按一定的生产计划来进行优化调度，这就涉及到显而易见的问题，即为了达到一定的计划产量该如何分配各原料的。

这个问题中的已知量是产量和原料质量，要求的是各原料用量。

在已得出的产量预测模型的基础上，我们改变模糊神经网络的输入输出，以产量、铁精矿硫份、焦炭灰份为网络输入，以铁精矿用量、焦炭用量为网络输出，得出生产计划模型。

利用已知数据对模型进行数据训练和验证的结果，如图 4.6 所示。

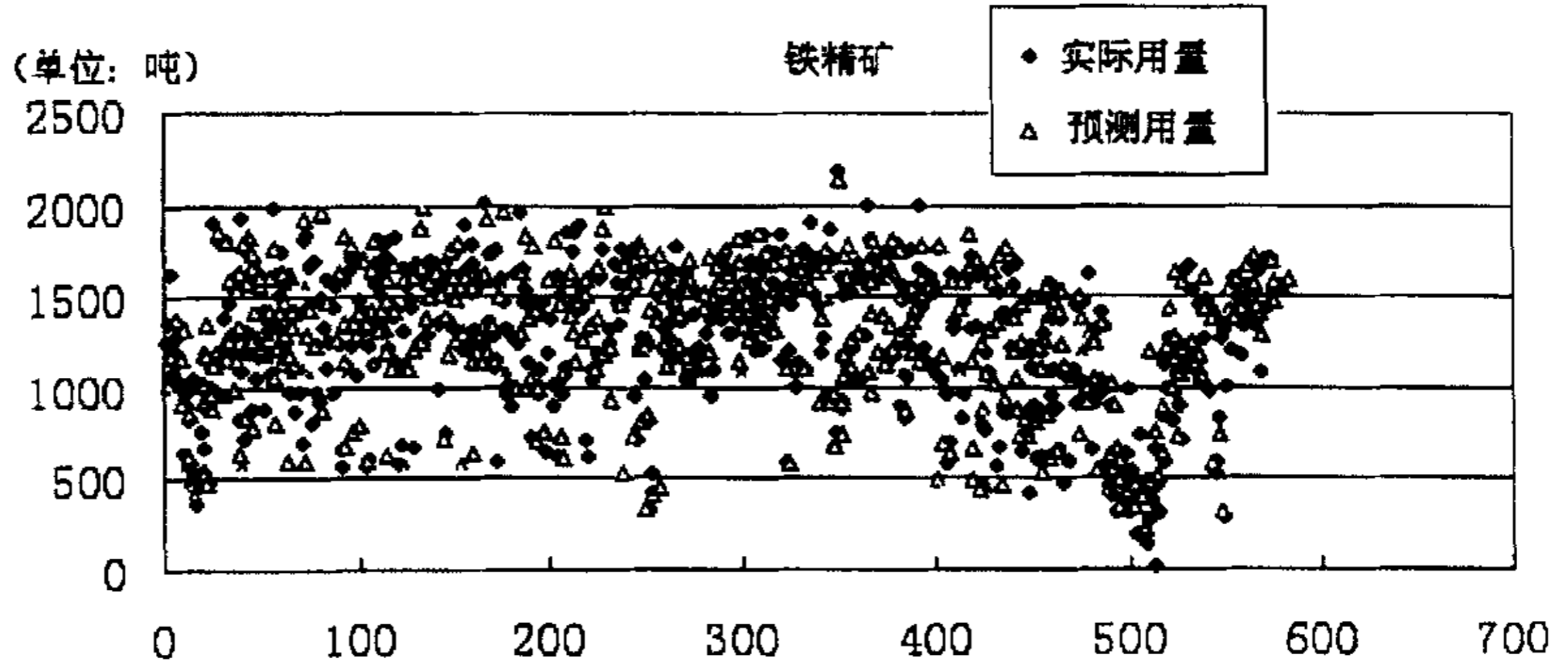


图 4.6 575 组预测铁精矿用量模糊神经网络输出与实际值的对比散点图

由图可见，对于现有数据的拟合较好，模型稳定性较强，将该模型用于调度系统，可以根据一定的计划产量，在已知铁精矿硫份和焦炭灰份的情况下，合理分配铁精矿和焦炭的用量，在保证产量的前提下节约原料。

4.5 本章小结

本章首先对熟球的综合质量进行评估，评定的结果为熟球质量较稳定，避免了创建熟球质量与产量相结合的新属性。随后利用神经网络挖掘产量预测模型。利用 BP 神经网络的学习能力来训练模型，并在最终的规则表达中运用模糊理论，避免了神经网络的过度拟合和不稳定性，得出适应性较强的产量预测模型。在产量预测模型的基础上，改变模糊神经网络的输入和输出，以产量、铁精矿硫份、焦炭灰份为网络输入，以铁精矿用量、焦炭用量为网络输出，得出生产计划模型，该模型拟合较好，可以根据计划产量，合理分配铁精矿和焦炭的用量，在保证产量的前提下节约原料。

5 全文总结

近年来,随着钢铁市场的国际化和国内钢铁生产竞争的加剧,球团工业面临着严峻的考验。为了在竞争中不被淘汰,必须改变原有的生产能耗物耗高,劳动生产率低,设备寿命短等一系列有碍企业发展的劣性。球团工业是典型的流程工业,建构一个综合决策、生产调度、车间管理、过程优化为一体的计算机集成生产自动化系统,是当前流程工业发展的方向。

本文运用回归统计方法和模糊神经网络技术给出了产量预测模型。

利用产量预测模型在已知铁精矿用量和原料质量的情况下,可以进行熟球产量的预测,以调整短期作业计划。

本课题围绕大冶球团厂的生产自动化系统中生产调度数据的挖掘应用,对基于线性回归的产量预测模型的挖掘过程作了深入探讨和研究,并利用模糊-神经网络优化模型,最终得出稳定性较好的产量预测模型,取得了若干具有一定价值的成果。

1. 本文在传统回归统计的基础上,通过聚类统计和相关性分析,利用决策树挖掘技术,将原料质量指标中对产量影响较大的属性加入产量预测模型,弥补了传统产量预测中因变量只包括原料用量的不足。

2. 本文利用神经网络挖掘产量预测和生产计划模型,并运用模糊理论,避免了神经网络的过度拟合和不稳定性,给出了有效而稳定的模型,在已知原料用量和质量的情况下,预测产量,调整短期作业计划,并根据一定的生产计划合理分配原料,在保证产量的前提下,节约原料。

3. 文中对球团厂一年的熟球质量统计数据进行标准化,通过统计分析得出关于每班各个等级的熟球在班总产量中的比例的分布,从而得出球团厂的熟球质量是较稳定的结论。该方法对于综合质量评估具有广泛的适用性。

致 谢

光阴似箭，三年的研究生生活即将过去，回首往事，我要向所有给予我关心和帮助的老师、同学、朋友及亲友们，致以诚挚的谢意！

在我参与的课题的研发工作中，始终得到了导师朱明富老师的悉心指导和热情关怀。在两年多的学习、工作与生活中，他为我倾注了大量的心血。他渊博的学识、敏捷的思维、严谨的治学态度和认真负责的工作作风，使我受益匪浅。他对我精心的培养和悉心的教导也将使我终身难忘。在此，我谨对朱老师表示最衷心的感谢。

再者感谢我们课题组王永骥教授、沈安文教授，郑定富老师。感谢他们一直以来在学习上、生活中、工作上给予我的关怀和帮助，特别是他们让我懂得做研究需要的一些优良品质和团体合作精神。

在课题的开发研究中，与同课题组的朱长华、刘珊、朱晓琳、方慧娟、白雷、邹明江、杨业等同学朝夕相处，使我感受到了集体的温暖，同学的友情和工作生活的乐趣。另外，师姐姚颖颖、苏静，师兄王保兵和好朋友关维也给了我无私、热情的帮助，在此一并向他们表示感谢。

我还要感谢系统工程研究所的全体领导、老师和同学，他们的支持、鼓励和帮助促使我努力工作，完成了学业。

最后，还要感谢我的父母，我的每一点进步都凝聚着他们的希望和心血，在任何困难面前，他们总是我最强大最有力最真挚的后盾与依靠，希望他们身体健康，万事顺心！

徐晓瑾

二零零四年 四月

于武昌喻园

参考文献

- [1] Fayyad, Piatetsky-shapiro G, Smyth. Chap. 1 From Data Mining to Knowledge Discovery An Overview Advances in Knowledge Discovery and Data Mining, 1996.
- [2] 陈京民等. 数据仓库与数据挖掘技术. 第 1 版. 北京: 电子工业出版社, 2002. 288~295
- [3] 陈玉明. 莱钢烧结厂的生产调度指挥系统. 烧结球团, 2001, 26 (6): 34~36
- [4] 谢维信等. 模糊神经网络研究. 深圳大学学报, 1999, 16 (2) :22~28
- [5] 任艳频, 张佐, 吴秋峰. 一类规则调度系统的 Petri 网研究方法. 计算机集成制造系统, 1999, 5 (2): 58~61
- [6] 李歧强. 生产过程的智能决策与调度:[博士学位论文]. 浙江杭州: 浙江大学控制系, 1998.
- [7] 杨家本. 第四讲连续过程 CIMS 中生产计划 /调度系统. 化工自动化及仪表, 1997, 24 (6): 52~56
- [8] 郑莉等. 现代统计学. 第 1 版. 北京: 中国纺织出版社, 2000. 5
- [9] 王军, 金以慧. 连续过程生产调度的研究策略. 系统工程理论与实践, 1998, (5): 40~46
- [10] 高俊波, 杨学兵, 蔡庆生. 一种基于分类的最佳工艺探索算法. 计算机应用, 2000, 20 (1): 37~39
- [11] Waltz E L, Buede D M.. Data fusion and decision support for command and control [J]. IEEE Transactions, Systems, Man, and Cybernetics, 1986, 16 (6) :865~879
- [12] 张长水. 用神经网络方法解决作业调度问题的研究. [博士学位论文]. 北京: 清华大学自动化系, 1992
- [13] David Hand, Heikki Mannila, Padhraic Smyth. Principles of Data Mining. 张银奎 廖丽 宋俊 等译. 北京: 机械工业出版社, 2003. 233~235

华中科技大学硕士学位论文

- [14] 宋擒豹, 蔡庆生. 神经网络神经挖掘方法中的数据准备问题. 计算机工程与应用, 2000, (12): 102~104
- [15] 熊熊, 张维. 数据挖掘技术及其实现. 预测, 2001, 20 (3): 64~68
- [16] 薛劲松, 宋宏等. CIMS 的总体设计. 第 1 版. 北京: 机械工业出版社, 1997. 244
- [17] C. J. Paul, L. E. Holloway, D. Yan extral. An Intelligent Reactive Monitoring and Scheduling System. IEEE Control Systems, June, 1992, 78~86
- [18] 李石君, 王汉飞, 周洞汝. 关系数据库中统计关系的挖掘和应用. 计算机工程与应用, 2000, (6): 117~118
- [19] Charly Kleissner. Data Mining for the Enterprise. IEEE, 1998
- [20] A. C. M. Fong, S. C. Hui, and G. Jha. Data Mining for Decision Support. IT Pro , Mar./Apr. 2002, pp. 9~17
- [21] 张佐, 谢东, 吴秋峰等. 一般生产调度问题的统一结构. [J]. 清华大学学报, 1999, 37 (4): 110~113
- [22] Jiawei Han, Micheline Kamber. Data Mining: Concepts and Techniques. 第 1 版. 范明, 孟小峰等译. 北京: 机械工业出版社, 2001.
- [23] 卢向南. 生产调度决策模型研究. 浙江大学学报 (人文科学版) 1999, 29 (6): 69~73
- [24] 李艳君, 吴铁军. 用于柔性流程工业生产调度的并行多目标遗传算法. 系统工程理论与实践, 2001, (6): 7~12
- [25] 常虹, 何丕廉. 神经网络与模糊技术的结合和发展. 计算机应用研究, 2001, (5): 4~6
- [26] 闻新, 周露等. MATLAB 神经网络应用设计. 北京: 科技出版设, 2000
- [27] 李晓峰, 刘光中. 人工神经网络 BP 算法的改进及其应用 [J]. 四川大学学报 (工程科学版), 2000, 32 (3): 105~109
- [28] 张汉江等. 模糊神经网络研究综述. 模糊系统与数学, 1998, 12 (1): 23~28
- [29] 刘俊强, 伞冶等. 模糊神经网络的快速学习算法研究. 高技术通讯, 2000 (6): 48~50
-

华中科技大学硕士学位论文

- [30] Lin CT, George Lee C S. Neural-network-based fuzzy logic control and decision system. IEEE Trans. Computers, 1991, 40(12):1320~1336
- [31] 陈荣秋等. 生产计划与控制. 第1版. 武汉: 华中理工大学出版社, 1995.7
- [32] Ming-Syan Chen. Data Mining: An Overview from a Database Perspective. IEEE Trans. Knowl Data Eng, 1996. 11;8(6):866~883
- [33] P Cheeseman J Stutz. Bayesian Classification(AutoClass): Theory and Results. Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996:153~180
- [34] Wen~Chi Hou. Extraction and Applications of Statistical Relationship in Relational Databases. IEEE Trans Knowl Data Eng, 1996. 11;8(6):939~945
- [35] 王致欧. 神经网络在系统辨识中的应用研究. 计算技术与自动化, 1999, Vol. 14(1): 15
- [36] 从爽. 面向 MATLAB 工具箱的神经网络理论与应用. 第一版. 合肥: 中国科学技术大学出版社, 1998, 11: 45~73
- [37] 龚菲. 基于神经网络的 PID 参数自整定方法研究. 保存地点: 华中科技大学控制系资料室, 2002
- [38] F. Alonso, J. L. Fuertes, L. Martez and C. Montes The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study, Pages 185~199
- [39] C. W. Kang and M. W. Golay
Neural-fuzzy modeling of plastic injection molding machine for intelligent control, Pages 33~43
- [40] 徐国祥, 胡清友. 统计预测和决策. 上海财经大学出版社, 1998.6
- [41] Lynn Ling X Li. A fuzzy adaptive method for intelligent control, Pages 43~48
- [42] Jungsoon Yoo and Sung Yoo. An initial comparison of a fuzzy neural classifier and a decision tree based classifier, Pages 375~381
-

- [43] Wang. Wen Jie, Tang. Bing Yong. A fuzzy adaptive method for intelligent control [J]. Expert systems with Applications, 1999, Vol.16(1):43~48
- [44] Chen. Chyi-Tsong, Peng. Shih-Tein. Intelligent process control using neural fuzzy techniques [J]. Journal of Process Control, 1999, Vol.9(6): 493~503
- [45] D. T. Pham, D. Karaboga. Self-tuning fuzzy controller design using genetic optimization and neural network modeling. Artificial Intelligence in Engineering 1999, Vol.13: 119~130
- [46] Yongji Wang, Gregor Fernholz, Sebastian Engell. Pseudo linear Neural Networks based Nonlinear System Modeling And Predictive Control. IFAC99
- [47] 张立明. 人工神经网络的模型及其应用[M]. 上海: 复旦大学出版社, 1995
- [48] 王悦. 人工神经网络在经济效益综合评价中的应用[J]. 北京广播电视大学学报, 2002, 3(3): 33~37
- [49] 熊熊, 汪德馨. 利用模糊神经网络进行神经挖掘的一种算法. 系统工程学报, 2000, 15(1): 32~37
- [50] 冯述虎. 基于人工神经网络的矿山经营绩效综合评价模型[J]. 中国煤炭经济学院学报, 2001, 15(3): 257~261
- [51] 刘增良. 模糊技术与神经网络技术. 北京航空航天大学出版社, 2001
- [52] 唐全. 专家系统中模糊概念的表达及应用. 南京邮电学院学报, 1997, 16(1): 54~57

附录 1 （攻读硕士学位期间发表的论文）

- [1] 徐晓瑾, 朱明富. 球团生产监控管理系统软件设计. 控制工程, 2004, 3

附录 2 （数据挖掘的各数据表）

1. 产量预测数据挖掘原数据表

铁精矿 用量 (单位: 吨)	焦炭 用量 (单位: 吨)	铁精矿 硫份 (单位: %)	焦炭 灰份 (单位: %)	熟球 产量 (单位: 吨)
1246	65	0.29	14.21	1100
1242	64	0.26	18.07	1184
1361	72	0.22	20.52	1186
1621	83	0.28	17.93	1010
1082	56	0.26	17.69	1129
1173	61	0.25	18.42	957
1280	61	0.31	20.85	1203
1033	48	0.33	21.55	847
1344	75	0.24	14.12	918
640	33	0.26	17.82	594
971	52	0.23	18.97	892
903	46	0.28	16.41	771
826	44	0.17	24.69	733
460	22	0.30	19.77	450
1056	57	0.27	14.65	954
361	18	0.30	19.30	110
2234	115	0.25	19.48	504
755	43	0.26	12.06	408
531	28	0.21	21.81	456
670	41	0.20	12.37	517
887	46	0.23	20.53	592
484	26	0.27	13.87	332
1116	59	0.24	18.36	775
1000	54	0.21	20.10	1066
1906	93	0.30	19.65	1350
1168	63	0.23	17.31	1000
1804	98	0.26	14.42	1394

铁精矿 用量 (单位: 吨)	焦炭 用量 (单位: 吨)	铁精矿 硫份 (单位: %)	焦炭 灰份 (单位: %)	熟球 产量 (单位: 吨)
1191	63	0.26	16.86	1150
2140	117	0.25	14.79	1403
1387	77	0.27	12.21	1251
1808	100	0.25	14.50	1432
1237	68	0.20	18.97	993
1461	79	0.22	18.86	1382
1284	68	0.27	15.95	1056
1540	85	0.21	17.39	1069
1164	61	0.25	18.57	875
1617	88	0.23	17.80	1241
824	45	0.22	17.97	611
1088	55	0.22	23.44	856
712	42	0.25	10.37	531
1195	61	0.27	18.22	1084
1307	68	0.26	18.07	1004
1740	85	0.30	19.19	1321
880	48	0.25	15.46	712
1655	90	0.21	19.86	1222
1045	54	0.26	17.72	910
1600	81	0.26	19.75	1062
1165	58	0.27	20.09	1003
1611	79	0.28	21.17	1370
878	44	0.31	17.01	599
1421	73	0.24	21.15	1237
1382	71	0.27	17.88	1013
1312	73	0.26	12.91	995
1076	58	0.24	16.75	1040

华中科技大学硕士学位论文

1984	97	0.34	15.69	1422
1340	76	0.25	12.01	1060
1616	92	0.21	15.95	1354
1501	75	0.32	15.36	972
1320	66	0.27	19.36	1237
1747	95	0.22	18.22	922
1567	77	0.28	19.85	1240
1246	61	0.31	18.05	959
965	53	0.23	16.76	526
1116	58	0.26	16.93	848
1101	57	0.26	18.07	796
865	46	0.22	20.69	621
413	20	0.30	19.62	234
968	54	0.26	13.13	571
270	14	0.28	17.56	275
698	38	0.23	16.53	642
1804	95	0.24	18.92	1052
1836	96	0.27	16.30	1395
1447	72	0.26	20.60	1085
1657	81	0.33	16.66	1277
801	45	0.25	12.80	572
1690	77	0.34	21.39	1241
974	47	0.31	18.52	907
1242	70	0.19	18.90	1166
917	54	0.20	15.12	852
1483	77	0.24	19.70	1278
1333	78	0.25	10.49	833
1600	80	0.26	20.49	1312
1101	56	0.29	17.20	870
1255	73	0.23	12.36	1167
970	55	0.23	13.51	781
966	51	0.25	17.82	650
1431	75	0.26	17.86	1114
1550	84	0.27	14.18	1395
1231	63	0.30	15.84	699
560	27	0.28	21.40	590
648	38	0.20	15.81	478

1576	80	0.29	17.30	1281
1111	55	0.29	19.60	880
1641	81	0.28	19.89	1085
1225	60	0.31	17.61	1006
1700	94	0.27	12.34	1322
1325	77	0.24	12.26	990
1070	56	0.25	18.49	921
1473	90	0.17	14.99	1086
1379	69	0.28	18.89	1150
1419	78	0.26	13.05	1114
1355	72	0.24	18.58	597
556	29	0.24	19.09	220
601	33	0.22	18.34	621
1218	67	0.26	14.16	935
1604	84	0.27	16.16	1357
1115	58	0.25	18.53	731
1677	96	0.24	12.43	1339
1560	84	0.26	15.10	1116
1806	100	0.25	13.97	1362
1504	79	0.23	20.21	991
1645	77	0.31	22.39	1163
1791	90	0.28	19.33	1175
1712	88	0.31	14.96	1373
1205	64	0.24	17.89	856
1447	81	0.22	16.53	1110
1139	60	0.29	14.64	768
1403	73	0.26	18.24	412
1822	99	0.23	17.53	1214
573	32	0.26	13.80	563
674	40	0.19	15.15	319
1304	66	0.30	16.37	946
1645	88	0.22	19.16	1262
1557	89	0.24	13.45	1042
1633	83	0.31	15.41	1221
1430	80	0.23	15.75	946
1480	76	0.25	20.15	1184
670	34	0.24	22.49	321

华中科技大学硕士学位论文

1670	78	0.32	21.48	1229
1612	81	0.26	19.94	918
1860	101	0.22	18.18	1207
1248	69	0.22	17.41	848
1235	64	0.32	12.70	980
1280	63	0.33	15.95	861
1610	79	0.32	17.14	1087
1290	71	0.21	17.89	907
1683	93	0.25	14.18	1173
1506	82	0.25	15.21	805
1680	85	0.27	18.87	1112
990	56	0.24	13.63	1021
1580	90	0.22	14.87	1097
1361	72	0.24	18.01	978
1637	77	0.30	21.80	1149
744	37	0.35	14.04	619
1669	85	0.27	19.08	1194
1366	75	0.25	14.81	1114
1551	80	0.29	16.26	1058
1327	73	0.24	15.39	920
1745	81	0.33	20.98	1322
1609	78	0.30	19.78	961
905	55	0.22	10.88	635
1217	67	0.24	16.12	897
1625	79	0.31	18.86	1208
1216	67	0.27	12.53	1036
1892	98	0.27	16.58	1158
1191	69	0.20	15.92	662
1234	63	0.27	18.68	1058
1677	88	0.28	14.59	1040
1780	96	0.24	17.44	1114
1315	74	0.22	15.64	917
1540	86	0.25	12.96	1081
1609	90	0.22	16.05	1085
1557	84	0.26	15.83	1167
1223	64	0.27	16.33	855
2013	101	0.31	16.30	1244

1262	70	0.27	12.43	918
1725	89	0.24	20.48	1158
1351	63	0.32	21.17	849
1723	88	0.29	16.82	1115
1276	68	0.27	14.99	905
591	30	0.28	17.84	421
1752	96	0.24	16.00	1303
1564	79	0.29	17.39	943
1590	89	0.27	11.93	1110
1458	70	0.28	22.26	960
1319	72	0.22	18.18	1188
952	46	0.32	19.35	589
1017	49	0.28	22.00	749
888	56	0.19	11.01	864
2251	119	0.28	14.97	1331
1259	62	0.31	17.73	1012
1622	81	0.28	18.78	980
1236	62	0.31	15.54	1004
1958	96	0.31	18.09	1159
1467	85	0.23	13.52	886
1644	97	0.18	16.63	1271
1532	75	0.32	16.29	981
1124	56	0.27	20.07	1124
714	38	0.26	15.60	714
1430	79	0.25	14.50	1032
1430	84	0.20	14.52	860
1426	73	0.27	18.04	1105
1396	81	0.21	14.96	977
1395	71	0.23	22.01	854
1084	56	0.25	18.92	600
1466	74	0.29	18.16	1148
641	33	0.27	16.75	124
242	13	0.21	18.15	398
1188	56	0.31	21.23	759
1371	79	0.26	10.49	990
890	41	0.33	21.08	623

华中科技大学硕士学位论文

2. 归一化数据, 用于神经网络训练和验证数据

铁精矿 用量	焦炭用量	铁精矿 硫份	焦炭灰份	熟球产量					
0.5452	0.5493	0.3011	0.6645	0.3260	0.6747	0.7175	0.4246	0.2740	0.4960
0.5434	0.5403	0.2927	0.5100	0.5329	0.5090	0.5120	0.2607	0.4375	0.5595
0.5959	0.6034	0.3077	0.3005	0.6637	0.7087	0.7383	0.4142	0.3305	0.5180
0.7104	0.6934	0.3633	0.5875	0.5254	0.3592	0.3794	0.2156	0.2885	0.5274
0.4729	0.4743	0.2431	0.4990	0.5126	0.4755	0.4626	0.2108	0.3160	0.8199
0.5130	0.5154	0.2623	0.4505	0.5513	0.3098	0.3524	0.2557	0.4275	0.1208
0.5602	0.5110	0.2207	0.7300	0.6812	0.5227	0.5162	0.2745	0.5315	0.5405
0.4513	0.3996	0.1601	0.8355	0.7185	0.5721	0.5729	0.2921	0.4790	0.5326
0.5884	0.6334	0.3683	0.4060	0.3216	0.7629	0.7150	0.3370	0.6915	0.5923
0.2781	0.2788	0.1510	0.5185	0.5191	0.3839	0.4019	0.2120	0.4560	0.3929
0.4240	0.4337	0.2271	0.3545	0.5806	0.7254	0.7541	0.4143	0.2270	0.6285
0.3940	0.3907	0.1954	0.6225	0.4437	0.4566	0.4576	0.2528	0.5020	0.5137
0.3601	0.3664	0.1911	0.0745	0.8864	0.7012	0.6806	0.3433	0.5130	0.6223
0.1988	0.1883	0.0856	0.6780	0.6234	0.5095	0.4883	0.2366	0.5570	0.6407
0.4614	0.4774	0.2717	0.5395	0.3496	0.7060	0.6612	0.3028	0.5900	0.6982
0.2243	0.2189	0.1074	0.5470	0.6163	0.3830	0.3675	0.1700	0.7300	0.4758
0.1551	0.1472	0.0665	0.7185	0.5984	0.6223	0.6097	0.2861	0.4015	0.6974
0.9806	0.9680	0.4729	0.4505	0.6080	0.6051	0.5973	0.3201	0.5470	0.5227
0.3288	0.3597	0.2266	0.4760	0.2113	0.5743	0.6141	0.3647	0.4975	0.2567
0.2301	0.2331	0.1143	0.2715	0.7327	0.4703	0.4902	0.2522	0.3955	0.4621
0.2913	0.3428	0.2480	0.2005	0.2278	0.8704	0.8145	0.3782	0.8845	0.4051
0.3870	0.3863	0.1877	0.3735	0.6641	0.5866	0.6428	0.4123	0.4535	0.2087
0.2093	0.2205	0.1289	0.5505	0.3081	0.7082	0.7741	0.4525	0.2465	0.4191
0.4879	0.4966	0.2601	0.4025	0.5479	0.6576	0.6283	0.3039	0.8185	0.3874
0.4368	0.4505	0.2391	0.2600	0.6409	0.5778	0.5582	0.2786	0.5585	0.6017
0.8361	0.7785	0.3598	0.6930	0.6174	0.7660	0.7996	0.4490	0.2985	0.5405
0.5108	0.5312	0.2969	0.3720	0.4920	0.6866	0.6505	0.3100	0.6115	0.6277
0.7911	0.8267	0.4520	0.4980	0.3375	0.5452	0.5149	0.2502	0.7285	0.5317
0.5209	0.5271	0.2754	0.5050	0.4681	0.4213	0.4460	0.2367	0.3385	0.4627
0.9392	0.9866	0.5412	0.4550	0.3573	0.4879	0.4911	0.2530	0.5245	0.4717
0.6073	0.6432	0.4148	0.5715	0.2190	0.4813	0.4831	0.2465	0.4750	0.5325
0.7929	0.8375	0.4667	0.4525	0.3415	0.3773	0.3836	0.1948	0.2905	0.6730
0.5412	0.5746	0.3327	0.2005	0.5805	0.1781	0.1680	0.0754	0.7130	0.6154
0.6399	0.6600	0.3565	0.3150	0.5750	0.4227	0.4518	0.2643	0.4970	0.2686
0.5619	0.5697	0.3038	0.5440	0.4195	0.1150	0.1158	0.0597	0.5965	0.5055
					0.3037	0.3210	0.1682	0.3710	0.4505
					0.7911	0.8009	0.4154	0.3820	0.5778

华中科技大学硕士学位论文

0.8052	0.8084	0.4224	0.5565	0.4379	0.7920	0.8393	0.4770	0.4685	0.3132
0.6338	0.6080	0.2953	0.5190	0.6681	0.6589	0.6652	0.3379	0.3275	0.6469
0.7263	0.6802	0.3076	0.8325	0.4573	0.7210	0.6444	0.2634	0.7325	0.7636
0.3491	0.3802	0.2318	0.4505	0.2507	0.7854	0.7540	0.3735	0.5795	0.5998
0.7409	0.6462	0.2551	0.8985	0.7099	0.7506	0.7354	0.3770	0.7305	0.3661
0.4253	0.3962	0.1845	0.7715	0.5568	0.5271	0.5385	0.2869	0.4105	0.5228
0.5434	0.5852	0.3524	0.1500	0.5768	0.6338	0.6780	0.3734	0.2985	0.4502
0.4002	0.4505	0.2820	0.1975	0.3748	0.4980	0.5015	0.2710	0.6455	0.3492
0.6496	0.6491	0.3229	0.3970	0.6199	0.6144	0.6094	0.3284	0.5075	0.5415
0.5835	0.6515	0.4613	0.4665	0.1272	0.7990	0.8279	0.4613	0.3640	0.5037
0.7012	0.6760	0.3332	0.5015	0.6622	0.2486	0.2658	0.1499	0.4840	0.3043
0.4813	0.4684	0.2236	0.6505	0.4858	0.2931	0.3364	0.2187	0.1385	0.3765
0.5491	0.6127	0.4018	0.3700	0.2271	0.5707	0.5548	0.2694	0.6950	0.4415
0.4235	0.4656	0.2848	0.3700	0.2890	0.7210	0.7385	0.3922	0.3235	0.5911
0.4218	0.4264	0.2210	0.4640	0.5194	0.6822	0.7468	0.4581	0.3760	0.2858
0.6267	0.6263	0.3196	0.4965	0.5215	0.7157	0.6957	0.3470	0.7410	0.3902
0.6792	0.7033	0.4090	0.5500	0.3247	0.6263	0.6707	0.3740	0.3365	0.4088
0.5386	0.5278	0.2632	0.6905	0.4134	0.6483	0.6396	0.3076	0.4255	0.6439
0.2428	0.2280	0.1010	0.6120	0.7106	0.2913	0.2821	0.1246	0.4090	0.7688
0.2816	0.3162	0.1925	0.1850	0.4120	0.7320	0.6527	0.2685	0.7915	0.7151
0.6906	0.6736	0.3566	0.6255	0.4913	0.7065	0.6839	0.3420	0.5145	0.6324
0.4857	0.4590	0.2166	0.6460	0.6142	0.8158	0.8502	0.4763	0.3055	0.5385
0.7193	0.6804	0.3233	0.6145	0.6300	0.5461	0.5778	0.3361	0.2975	0.4973
0.5359	0.5081	0.2512	0.7360	0.5082	0.5403	0.5356	0.3003	0.8125	0.2455
0.7453	0.7894	0.5083	0.5590	0.2263	0.5602	0.5281	0.2449	0.8485	0.4190
0.5800	0.6462	0.4243	0.3790	0.2219	0.7056	0.6627	0.2981	0.7945	0.4827
0.4676	0.4672	0.2341	0.4730	0.5549	0.5646	0.5969	0.3449	0.2740	0.5227
0.6452	0.7538	0.5281	0.0590	0.3678	0.7378	0.7812	0.4405	0.4615	0.3244
0.6038	0.5830	0.2928	0.5850	0.5767	0.6598	0.6887	0.3678	0.4640	0.3795
0.6214	0.6589	0.4144	0.5210	0.2640	0.7364	0.7154	0.3657	0.5555	0.5756
0.5932	0.6025	0.3145	0.3935	0.5599	0.4323	0.4710	0.2820	0.3970	0.2954
0.2411	0.2463	0.1254	0.3820	0.5871	0.6924	0.7529	0.4414	0.3240	0.3617
0.2609	0.2741	0.1511	0.3055	0.5470	0.5959	0.6059	0.3196	0.4195	0.5296
0.5328	0.5611	0.3114	0.4915	0.3235	0.7175	0.6511	0.2765	0.6950	0.7323
0.7030	0.7055	0.3685	0.5685	0.4304	0.3239	0.3087	0.1523	0.9270	0.3173
0.4874	0.4904	0.2497	0.4405	0.5570	0.7316	0.7113	0.3631	0.5415	0.5865
0.7351	0.8090	0.5176	0.4100	0.2310	0.5981	0.6301	0.3451	0.4520	0.3581
0.6836	0.7091	0.3752	0.4935	0.3736	0.6796	0.6717	0.3391	0.6265	0.4358

华中科技大学硕士学位论文

0.5809	0.6134	0.3333	0.4120	0.3895	0.6386	0.5874	0.2555	0.6125	0.7567
0.7651	0.6796	0.2797	0.8350	0.6881	0.5773	0.6020	0.3354	0.3115	0.5386
0.7052	0.6553	0.3002	0.6990	0.6238	0.4156	0.3830	0.1716	0.7765	0.6010
0.3949	0.4633	0.3523	0.2790	0.1483	0.4442	0.4143	0.1845	0.5780	0.7426
0.5324	0.5616	0.3001	0.3795	0.4282	0.3874	0.4704	0.3845	0.1550	0.1549
0.7122	0.6628	0.3095	0.7415	0.5749	0.9881	1.0019	0.5504	0.5845	0.3668
0.5320	0.5616	0.3549	0.5700	0.2364	0.5509	0.5222	0.2576	0.7295	0.5144
0.8299	0.8270	0.4236	0.5710	0.4529	0.7109	0.6819	0.3392	0.6150	0.5708
0.5209	0.5791	0.3508	0.1925	0.4178	0.5408	0.5247	0.2586	0.7485	0.3974
0.5399	0.5286	0.2740	0.5410	0.5653	0.8590	0.8070	0.3910	0.7375	0.5339
0.7351	0.7430	0.4078	0.6235	0.3464	0.6426	0.7105	0.4439	0.3365	0.2893
0.7805	0.8054	0.4452	0.3855	0.4989	0.7206	0.8147	0.5190	0.0850	0.4557
0.5756	0.6206	0.3513	0.3195	0.4024	0.6712	0.6332	0.2924	0.8220	0.4377
0.6747	0.7263	0.4371	0.4665	0.2593	0.4914	0.4698	0.2261	0.5705	0.6396
0.7052	0.7567	0.4230	0.3105	0.4244	0.3107	0.3215	0.1773	0.5035	0.4005
0.6822	0.7019	0.3597	0.4870	0.4126	0.6263	0.6668	0.3760	0.4270	0.3415
0.5350	0.5391	0.2815	0.5505	0.4393	0.6263	0.7038	0.4470	0.2235	0.3429
0.8832	0.8467	0.4042	0.7470	0.4378	0.6245	0.6162	0.3292	0.5400	0.5308
0.5522	0.5893	0.3820	0.5325	0.2312	0.6113	0.6772	0.4128	0.2525	0.3664
0.7563	0.7464	0.3593	0.4025	0.6617	0.6108	0.5994	0.2812	0.3490	0.7433
0.5914	0.5297	0.2198	0.7940	0.6985	0.4738	0.4718	0.2333	0.4625	0.5782
0.7554	0.7369	0.3591	0.6485	0.4658	0.6421	0.6181	0.3126	0.6360	0.5375
0.5584	0.5739	0.3211	0.5415	0.3678	0.2785	0.2798	0.1411	0.5675	0.4618
0.2565	0.2511	0.1286	0.6210	0.5204	0.1027	0.1120	0.0629	0.2580	0.5368
0.7682	0.8038	0.4260	0.4085	0.4218	0.5196	0.4683	0.1960	0.7665	0.7017
0.6853	0.6631	0.3446	0.6540	0.4965	0.6003	0.6629	0.4975	0.5060	0.1271
0.6968	0.7447	0.4949	0.5485	0.2041	0.3883	0.3461	0.1408	0.8365	0.6935

华中科技大学硕士学位论文

3. 球团厂一年内的熟球抗压强度质量指标的标准化数据（用于综合质量评估）

一级	二级	三级	等外
0.54	0.19	0.1	0.17
0.57	0.15	0.12	0.16
0.52	0.16	0.14	0.18
0.54	0.12	0.11	0.23
0.55	0.16	0.12	0.17
0.55	0.15	0.15	0.15
0.52	0.13	0.17	0.18
0.53	0.16	0.13	0.18
0.54	0.11	0.13	0.22
0.58	0.17	0.15	0.10
0.55	0.16	0.07	0.22
0.54	0.18	0.15	0.13
0.56	0.16	0.18	0.10
0.52	0.14	0.16	0.18
0.53	0.12	0.14	0.21
0.54	0.12	0.16	0.18
0.59	0.1	0.14	0.17
0.56	0.17	0.14	0.13
0.54	0.24	0.15	0.07
0.55	0.15	0.15	0.15
0.55	0.17	0.17	0.11
0.53	0.17	0.17	0.13
0.54	0.14	0.18	0.14
0.58	0.15	0.13	0.14
0.53	0.08	0.09	0.30
0.52	0.15	0.18	0.15
0.57	0.02	0.1	0.31
0.57	0.08	0.17	0.18
0.55	0.14	0.14	0.17
0.55	0.16	0.17	0.12
0.52	0.12	0.16	0.20
0.56	0.15	0.14	0.15
0.5	0.16	0.17	0.17

一级	二级	三级	等外
0.57	0.14	0.11	0.18
0.56	0.13	0.17	0.14
0.54	0.1	0.12	0.24
0.55	0.09	0.12	0.24
0.55	0.16	0.12	0.17
0.56	0.13	0.15	0.16
0.56	0.13	0.15	0.16
0.55	0.11	0.09	0.25
0.56	0.11	0.15	0.18
0.54	0.13	0.11	0.22
0.54	0.06	0.1	0.30
0.54	0.13	0.14	0.19
0.56	0.13	0.14	0.17
0.53	0.15	0.14	0.18
0.57	0.15	0.1	0.18
0.53	0.14	0.11	0.22
0.55	0.17	0.13	0.15
0.54	0.18	0.19	0.09
0.56	0.17	0.11	0.16
0.55	0.15	0.15	0.15
0.56	0.12	0.16	0.16
0.54	0.11	0.17	0.18
0.58	0.18	0.16	0.08
0.55	0.15	0.14	0.16
0.55	0.12	0.1	0.23
0.57	0.14	0.18	0.11
0.55	0.13	0.12	0.20
0.54	0.18	0.12	0.16
0.54	0.14	0.13	0.19
0.57	0.14	0.15	0.14
0.52	0.13	0.13	0.22
0.54	0.13	0.17	0.16
0.53	0.14	0.13	0.20

华中科技大学硕士学位论文

0.51	0.15	0.12	0.22
0.56	0.12	0.15	0.17
0.57	0.1	0.18	0.15
0.57	0.12	0.14	0.17
0.53	0.19	0.13	0.15
0.51	0.17	0.12	0.20
0.55	0.17	0.16	0.12
0.54	0.13	0.13	0.20
0.57	0.13	0.1	0.20
0.57	0.09	0.16	0.18
0.6	0.11	0.14	0.15
0.58	0.17	0.17	0.08
0.52	0.11	0.14	0.23
0.54	0.11	0.15	0.20
0.56	0.13	0.15	0.16
0.58	0.14	0.18	0.10
0.54	0.14	0.15	0.17
0.52	0.07	0.16	0.25
0.54	0.15	0.15	0.16
0.51	0.07	0.14	0.28
0.58	0.16	0.15	0.11
0.58	0.11	0.13	0.18
0.56	0.17	0.14	0.13
0.55	0.15	0.16	0.14
0.55	0.17	0.12	0.16
0.49	0.12	0.16	0.23
0.54	0.19	0.12	0.15
0.57	0.1	0.13	0.20
0.56	0.09	0.11	0.24
0.51	0.15	0.13	0.21
0.54	0.12	0.17	0.17
0.55	0.05	0.17	0.23
0.55	0.17	0.13	0.15
0.58	0.13	0.13	0.16
0.52	0.1	0.11	0.27
0.53	0.21	0.13	0.13
0.54	0.13	0.13	0.20

0.55	0.18	0.14	0.13
0.52	0.15	0.13	0.20
0.55	0.12	0.14	0.19
0.5	0.16	0.14	0.20
0.53	0.19	0.13	0.15
0.55	0.15	0.14	0.16
0.47	0.15	0.17	0.21
0.56	0.17	0.16	0.11
0.54	0.11	0.12	0.23
0.56	0.08	0.16	0.20
0.55	0.14	0.14	0.17
0.54	0.16	0.14	0.16
0.55	0.08	0.12	0.25
0.54	0.12	0.13	0.21
0.49	0.14	0.12	0.25
0.53	0.11	0.16	0.20
0.54	0.13	0.1	0.23
0.54	0.16	0.19	0.11
0.53	0.14	0.15	0.18
0.57	0.16	0.15	0.12
0.53	0.16	0.14	0.17
0.54	0.15	0.17	0.14
0.53	0.13	0.18	0.16
0.57	0.15	0.12	0.16
0.59	0.15	0.16	0.10
0.56	0.11	0.15	0.18
0.56	0.19	0.13	0.12
0.54	0.11	0.12	0.23
0.56	0.13	0.12	0.19
0.52	0.14	0.13	0.21
0.53	0.15	0.18	0.14
0.56	0.19	0.1	0.15
0.55	0.16	0.12	0.17
0.55	0.14	0.14	0.17
0.54	0.17	0.14	0.15
0.5	0.17	0.16	0.17
0.51	0.14	0.15	0.20

华中科技大学硕士学位论文

0.55	0.14	0.14	0.17
0.56	0.1	0.14	0.20
0.51	0.17	0.17	0.15
0.5	0.13	0.13	0.24
0.56	0.18	0.12	0.14
0.54	0.16	0.11	0.19
0.54	0.18	0.16	0.12
0.55	0.17	0.16	0.12
0.55	0.13	0.14	0.18
0.55	0.17	0.15	0.13
0.55	0.08	0.12	0.25
0.56	0.16	0.15	0.13
0.55	0.15	0.17	0.13
0.57	0.2	0.14	0.09
0.56	0.13	0.12	0.19
0.56	0.19	0.13	0.12
0.55	0.15	0.15	0.15
0.56	0.12	0.14	0.18
0.54	0.12	0.13	0.21
0.55	0.11	0.15	0.19
0.53	0.14	0.13	0.20
0.57	0.19	0.15	0.09
0.55	0.11	0.14	0.20
0.58	0.08	0.11	0.23
0.56	0.12	0.16	0.16
0.53	0.17	0.13	0.17
0.53	0.11	0.17	0.19
0.55	0.13	0.13	0.19
0.55	0.13	0.14	0.18
0.56	0.15	0.12	0.17
0.51	0.17	0.09	0.23

0.57	0.12	0.12	0.19
0.54	0.12	0.11	0.23
0.54	0.09	0.16	0.21
0.56	0.09	0.15	0.20
0.54	0.16	0.13	0.17
0.57	0.16	0.18	0.09
0.55	0.05	0.19	0.21
0.56	0.1	0.13	0.21
0.53	0.12	0.15	0.20
0.55	0.15	0.13	0.17
0.55	0.2	0.14	0.11
0.57	0.16	0.12	0.15
0.58	0.13	0.16	0.13
0.55	0.15	0.14	0.16
0.55	0.14	0.17	0.14
0.52	0.14	0.18	0.16
0.53	0.11	0.12	0.24
0.55	0.14	0.17	0.14
0.54	0.13	0.15	0.18
0.53	0.18	0.17	0.12
0.53	0.1	0.19	0.18
0.56	0.16	0.19	0.09
0.53	0.04	0.15	0.28
0.54	0.15	0.06	0.25
0.52	0.16	0.11	0.21
0.53	0.11	0.15	0.21
0.56	0.17	0.14	0.13
0.54	0.14	0.12	0.20
0.53	0.11	0.11	0.25