

中图分类号:

单位代号: 10280

密 级:

学 号: 13720104

---

上海大学



硕士学位论文

---

SHANGHAI UNIVERSITY  
MASTER'S DISSERTATION

题 目	基于人工智能算法的高通量 Rietveld 精 修自动化程序: AutoFP 的设计与应用
--------	--

作 者 崔晓鹏

学科专业 凝聚态物理

导 师 张金仓 教授

完成日期 2016.5

姓 名:崔晓鹏

学号:13720104

论文题目: 基于人工智能算法的高通量 Rietveld 精修自动化程序: AutoFP 的设计与应用

## 上海大学

本论文经答辩委员会全体委员审查, 确认符合上海大学硕士学位论文质量要求。

答辩委员会签名:

主任:

委员:

导 师:

答辩日期:

## 原创性声明

本人声明：所呈交的论文是本人在导师指导下进行的研究工作。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发表或撰写过的研究成果。参与同一工作的其他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了谢意。

签 名：\_\_\_\_\_日期：\_\_\_\_\_

## 本论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定，即：学校有权保留论文及送交论文复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容。

（保密的论文在解密后应遵守此规定）

签 名：\_\_\_\_\_导师签名：\_\_\_\_\_日期：\_\_\_\_\_

---

上海大学理学硕士学位论文

**基于人工智能算法的高通量 Rietveld 精修  
自动化程序：AutoFP 的设计与应用**

姓 名：崔晓鹏

导 师：张金仓 教授

学科专业：凝聚态物理

上海大学理学院

2016 年 05 月

A Dissertation Submitted to Shanghai University for the  
Degree of Master in Science

**Design and Application of AutoFP:  
A Program for High-Throughput  
and Automated Rietveld Refinement  
Based on AI Algorithm**

MA Candidate: Xiaopeng Cui

Supervisor: Prof. Jincang Zhang

Major: Condensed Matter Physics

**Science College, Shanghai University**

**05, 2016**

# 摘 要

晶体结构和磁结构对固体材料的物理化学等属性具有决定性的作用。在研究晶体结构和磁结构时，单晶是最好的选择，但是单晶生长是非常困难的，有时候几乎是不可能的，甚至很多材料本来就不是单晶。因此粉末衍射成为了研究晶体结构和磁结构的强有力手段。但是在 Rietveld 方法之前，由于粉末衍射中存在着强烈的峰重叠，粉末衍射晶体学很难通过衍射图谱获取完整的结构信息。Rietveld 精修是一种通过利用最小二乘法拟合实验图谱和计算图谱分析粉末衍射数据的一种方法。这种方法通过利用衍射图谱全部信息，而不止是峰位置信息，不仅可以完美解决峰重叠难题，更可以帮助人们从粉末衍射中获取更多其他信息。由此，Rietveld 精修成为了粉末衍射数据分析的重要方法，但传统 Rietveld 精修程序普遍存在着效率低，操作复杂的缺点。本文主要研究了基于人工智能算法的自动化和高通量的 Rietveld 精修程序 AutoFP 的设计和应用。

第一章阐述了课题研究的目的和意义，以及国内外研究的现状。随着人工智能（AI）技术的发展，越来越多的复杂问题已经可以由人工智能算法加以解决。把人工智能算法引入到 Rietveld 精修领域，将有可能给 Rietveld 精修自动化带来新的思路。而 Rietveld 精修自动化的实现将为高通量 Rietveld 精修奠定基础。

第二章基于人工智能算法和广泛使用的 Rietveld 精修程序 FullProf 开发了一款高度自动化的 Rietveld 精修工具集 AutoFP。AutoFP 采用 Python 代码编写，开源免费，用户界面友好，容易学习和使用，具有良好的跨平台性。AutoFP 的开发极大地提高了 Rietveld 精修程序 FullProf 的自动化水平。它采用专家系统算法控制 FullProf 去模拟人类专家的精修过程，从而高度自动化，智能化地完成复杂的精修任务。由此证明了专家系统是实现 Rietveld 精修自动化的很好选择。同时，AutoFP 在专家系统的基础上提供了服务于机器学习算法的 Rietveld 精修策略搜索接口，并且利用 Q 学习算法来进行精修策略搜索来提升专家系统策略库。同时，AutoFP 也为高级用户保留了编程接口，方便了他们设计自己的

精修策略和精修算法，或者更换新的精修引擎。AutoFP 也可以被其他程序作为一个自动化的精修引擎所调用。

第三章研究了基于 AutoFP 的自动化定量相分析的实现。Rietveld 方法的定量相分析是粉末衍射定量相分析的重要手段，但手动的定量相分析往往复杂而低效。基于 AutoFP 的 Rietveld 精修自动化能力，本文继续开发了基于 AutoFP 的自动化定量相分析程序，大大简化了采用 FullProf 进行定量相分析的操作。

第四章研究了基于 AutoFP 的高通量 Rietveld 精修系统的设计。随着材料基因组计划的提出，高通量的材料表征成为重点，其中，高通量粉末衍射将产生大量的衍射数据，面对如此多的数据，高通量的衍射数据分析就变得非常迫切，本文基于自动化 Rietveld 精修程序 AutoFP 和上海大学自强 4000 集群 MPI 并行化框架开发了高通量的 Rietveld 精修系统，为高通量 Rietveld 精修创造了可能性，有助于材料基因组工程的推进。

第五章对程序进行了综合应用测试。测试表明 AutoFP 能够自动化，高质量地完成单相材料的 Rietveld 精修任务和混合物的定量相分析任务。基于 AutoFP 的高通量 Rietveld 精修系统能够完成海量粉末衍射数据的高通量 Rietveld 精修。

第六章对全文工作做了总结，并对该领域未来的发展做了展望。

**关键词：** AutoFP，Rietveld 精修，自动化，人工智能，高通量

## ABSTRACT

The crystal structure and magnetic structure is extremely important in determining the behavior and properties of a solid material. In the research of crystal structure and magnetic structure, a single crystal is the best choice. But single crystal growth is not easy sometime impossible and many materials are not single crystals. So powder diffraction becomes the powerful tool for the research of crystal structure and magnetic structure. Powder diffraction crystallography before Rietveld refinement is difficult to achieve the complete information of a crystal because there are strongly overlapping reflections. Rietveld refinement is a method of analyzing powder diffraction data in which the crystal structure and magnetic structure are refined by fitting the entire profile of the diffraction pattern to a calculated profile using a least-squares approach. Rietveld refinement uses all the data of powder diffraction not only the peak position. So it can not only solve the overlapping peaks challenge perfectly, but also get more information from the powder diffraction. So Rietveld refinement becomes an important method of powder diffraction data analysis. But traditional Rietveld refinement programs often have low efficiency and complex operation. The main attention of this article is paid to the design and application of AutoFP: a program for high-throughput and automated Rietveld refinement based on Artificial Intelligence (AI) algorithm.

In chapter 1, we elaborate the topic goal and significance of the research, as well as the research status at home and abroad. With the advancement of the AI technology, more and more complex problems can be solved by AI algorithms. Introducing AI algorithms to the field of Rietveld refinement will bring new thinking to realizing the automation of Rietveld refinement. The realization of the automation will lay the foundation for high-throughput Rietveld refinement.

In chapter 2, we use widely used Rietveld refinement program FullProf and AI



algorithms to developed a highly automated Rietveld refinement program AutoFP. AutoFP is an open-source software package developed in Python, and is user-friendly, easy to learn and use, has good cross-platform. AutoFP improves the extent of automation of the widely used Rietveld refinement program FullProf greatly. The expert system algorithm is used as the control algorithm to simulate the manual refinement process when FullProf program is used to do Rietveld refinement. This enables AutoFP complete the Rietveld refinement highly automatically and intelligently. It is proven that expert system algorithm is a good choice for automating Rietveld refinement. Meanwhile, AutoFP provides a machine learning algorithms interface and a Q-learning module for searching the strategies of Rietveld refinement to improve the policy library of the expert system. Programming interface is reserved for advanced users to implement their own refinement strategy or use new Rietveld refinement engines. AutoFP can also be used as an automated Rietveld refinement engine by other programs.

In chapter 3, we focus on the realization of automated quantitative phase analysis based on AutoFP. Rietveld quantitative phase analysis is an important means of powder diffraction quantitative phase analysis. But manual Rietveld quantitative phase analysis is often complex and inefficient. Based on AutoFP, we continued to develop an automated quantitative phase analysis program, which greatly simplify the operation of Rietveld quantitative phase analysis in FullProf.

In chapter 4, we focus on the design of high-throughput Rietveld refinement system. With the proposal of Materials Genome Initiative (MGI), high-throughput characterization of materials becomes the focus. High-throughput powder diffraction will produce large amounts of powder diffraction data. Faced with so much data, high-throughput powder diffraction analysis of high-throughput data becomes very urgent. Based on the automated Rietveld refinement program AutoFP and the MPI(Message Passing Interface) parallel framework of Shanghai University ZQ4000 cluster, we developed a high-throughput Rietveld refinement system, which will

create possibilities for the high-throughput Rietveld refinement and boost Materials Genome Initiative.

In chapter 5, we report the comprehensive application and testing of these programs. The results show that AutoFP can complete automated Rietveld refinement of single phase materials and the quantitative phase analysis of the mixture with high quality. The high-throughput Rietveld refinement system can complete the high-throughput Rietveld refinement task of massive powder diffraction data.

In chapter 6, we summarize the research and make prospects for future development in this field.

**Keywords:** AutoFP, Rietveld refinement, automated, AI, high-throughput

## 目 录

第一章 绪论.....	1
1.1 课题研究的目的和意义 .....	1
1.2 国内外研究概况.....	5
1.2.1 国外研究概况 .....	5
1.2.2 国内研究概况 .....	6
1.3 论文的主要研究内容 .....	6
第二章 基于人工智能算法的自动化 Rietveld 精修程序 AutoFP.....	7
2.1 AutoFP 程序架构.....	7
2.1.1 程序架构 .....	7
2.1.2 跨平台性 .....	8
2.1.3 设计原则 .....	10
2.1.4 使用说明 .....	11
2.1.5 附加工具集 .....	12
2.2 专家系统.....	13
2.2.1 专家系统的原理 .....	13
2.2.2 专家系统在 AutoFP 里的实现.....	15
2.2.3 精修策略 .....	18
2.3 基于人工智能算法的 Rietveld 精修策略搜索 .....	21
2.3.1 Q 学习算法搜索原理.....	21
2.3.2 Q 学习算法精修策略搜索的实现.....	22
第三章 基于 AutoFP 的自动化定量相分析 .....	25
3.1 FullProf 定量相分析 .....	25
3.1.1 Rietveld 定量相分析原理 .....	25
3.1.2 FullProf 定量相分析功能.....	26
3.2 AutoFP 自动化定量相分析.....	26
3.2.1 基于 AutoFP 自动化定量相分析的实现.....	26

3.2.2 AutoFP 自动化定量相分析的使用.....	27
<b>第四章 基于 AutoFP 的高通量 Rietveld 精修系统.....</b>	<b>28</b>
4.1 高通量 Rietveld 精修的提出.....	28
4.2 利用 AutoFP 实现高通量 Rietveld 精修 .....	29
4.2.1 系统架构.....	29
4.2.2 AutoFP 程序 MPI 并行化.....	31
4.2.3 高通量任务格式.....	32
4.2.4 数据交换格式.....	33
<b>第五章 AutoFP 的综合应用测试实例 .....</b>	<b>36</b>
5.1 测试说明 .....	36
5.2 AutoFP 单相 Rietveld 精修测试 .....	36
5.3 基于 AutoFP 的自动化定量相分析测试 .....	39
5.4 基于 AutoFP 的高通量 Rietveld 精修系统测试.....	40
<b>第六章 结论与展望.....</b>	<b>43</b>
6.1 结论.....	43
6.2 展望.....	44
<b>参考文献.....</b>	<b>45</b>
<b>作者在攻读硕士学位期间公开发表的论文.....</b>	<b>51</b>
<b>作者在攻读硕士学位期间所承担的项目 .....</b>	<b>52</b>
<b>致    谢.....</b>	<b>53</b>

# 第一章 绪论

## 1.1 课题研究的目的是和意义

课题研究的主要内容是开发一款自动化和高通量的 Rietveld 精修程序。目前, Rietveld 精修程序的自动化水平都比较低[1-2], 需要大量的人为干涉, 操作复杂, 费时费力, 也难以满足高通量 Rietveld 精修的需求。随着人工智能(AI)技术的发展, 越来越多的复杂问题已经可以由人工智能算法加以解决, 比如, 2016年基于深度神经网络人工智能算法的谷歌围棋程序 AlphaGo 连续击败多位世界围棋冠军[3]。如果把人工智能算法这个在机器人领域应用广泛的算法应用到需要很多人干涉的 Rietveld 精修进程中, 将有可能给解决复杂的晶体学精修问题带来新的思路, 这个想法的实现将为 Rietveld 精修自动化带来重要的进步, 并且将有力推动高通量材料分析技术的发展, 进而推动材料基因组工程的实施。

Rietveld 精修[4]是基于粉末衍射的晶体结构和磁结构[5-7] (中子衍射)测定流程中的最后一个重要的步骤, 其在粉末衍射结构分析中具有至关重要的地位。晶体结构和磁结构测定是凝聚态物理, 材料物理以及化学化工等研究的重要基础, 对任何新物质的认识都是从认识其结构开始的。晶体结构和磁结构测定从样品结晶形态来分可以分为单晶和粉晶结构测定技术[1,8]。单晶衍射晶体结构测定的衍射技术和算法软件发展的相对比较成熟, 但并不是所有研究对象都可以生长出符合结构测定要求的单晶体, 甚至有些材料根本无法以单晶形式存在。实验中最先得到的材料大都是多晶(粉末)材料, 再加上粉末衍射实验(X射线粉末衍射, 中子粉末衍射)简单快捷, 吸引着众多研究者发展粉末衍射晶体结构测定技术, 并在近年来成为凝聚态物理、材料科学技术和晶体学界的一个非常活跃的研究课题[9-11]。粉末衍射测定晶体结构和磁结构, 实验比较容易实现, 但由于粉末衍射数据是把三维数据压缩到一维, 造成一定的信息丢失以及择优取向, 峰重叠等问题[8], 因此用粉末衍射法进行结构分析至今仍不是一件容易的工作[2]。粉末衍射法测定晶体结构需要多个步骤, 多个软件的配合使用才

能最终完成[2,9]。图 1 - 1 给出了通常粉末衍射法测定晶体结构的步骤框图[12]，从图中看出，Rietveld 结构精修是精确测定原子坐标排列的最后一步，Rietveld 精修通常是利用最小二乘算法来对实验衍射谱进行全谱拟合来获取晶体信息[13]。而对 Rietveld 精修的收敛起决定作用的是初始结构模型的正确与否[14]。如果初始结构模型比较接近真实结构模型时，Rietveld 精修就能收敛到符合实际需求的精度。在坐标空间搜索结构模型的步骤中已经有基于遗传算法和粒子群法算法等新型算法的尝试[15-17]。

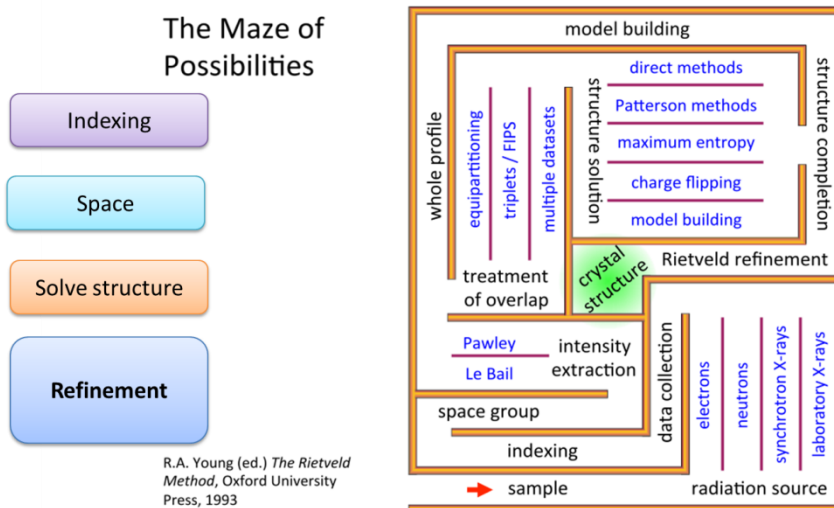


图 1 - 1 Rietveld 精修在晶体结构求解中的位置

提高 Rietveld 精修的自动化水平，具有重要的应用价值。晶体结构的 Rietveld 精修有两类主要用途。第一类是对已知材料进行掺杂取代后，晶体中的原子位置或原子磁矩排列发生了变化，通过 Rietveld 精修可以得到原子位置或原子磁矩排列的细微变化，从而获得精确的结构信息，对已知材料的新功能开发具有重要意义。第二类是对于新型材料，其晶体结构和磁结构是完全未知的(比如 80 年代末新开发的  $YBa_2Cu_3O_7$  高温超导材料晶体结构的精确测定花费了晶体学家将近一年时间[18])，精确测定其晶体结构或磁结构是研究该新型材料物理性能的基础，对未知的晶体进行 Rietveld 精修是新型功能材料开发必须进行的研究工作。因此，提高 Rietveld 精修的自动化水平具有重要的应用价值。

提高 Rietveld 精修自动化水平, 有利于提高新材料的研发速度。Rietveld 精修目前的主要使用的程序有 DBWS[19], FullProf[20], GSAS[21-22], Rietan[23] 等, DBWS[19]程序是出现较早的一个 Rietveld 精修程序, 其使用一个控制文件来控制精修过程, 但其控制文件复杂难懂, 需要花费大量时间学习; FullProf[20]程序是基于 DBWS 的内核由法国人开发的, 其把控制文件进行了一定程度的改进, 也开发了界面程序, 方便了程序的使用, 但也仅仅是把控制文件变成了界面程序, 并没有提高 Rietveld 精修的自动化程度; GSAS 是美国人开发的一个 Rietveld 精修程序, 使用命令行交互式精修策略, 参数众多, 操作复杂, 虽然后来有人开发了 GSAS 的界面程序 EXPGUI, 但并没有改变 GSAS 的交互操作本质[21,22], Rietan[23]程序是日本人开发, 也是基于控制文件的。这些程序目前的自动化水平都较低, 需要大量的人为干涉, 基本上都需要具有较深的晶体学知识才能很好的使用[2]。而对于新材料开发人员来说, 将不得不花费大量的时间来学习晶体学知识来使用这些程序, 使新材的开发和新材料晶体结构的测定时间大大变长, 延迟了新材料开发的进度。因此, 提高 Rietveld 精修的自动化水平, 减少程序运行中的人为干涉, 减少使用人员的学习时间, 有利于提高新材料的开发速度。

Rietveld 精修自动化和高通量程序的实现, 有利用提升我国在晶体学软件开发方面的水平, 使之与我国硬件水平的进步相匹配。晶体结构和磁结构测定一般采用衍射方法, 结构测定技术的发展进步主要包括硬件和软件两个方面, 硬件方面是指衍射装置的技术进步, 软件方面是指晶体结构测定理论、算法和软件技术的进步。在硬件技术发展方面, 商用高精度的 X 射线衍射仪在我国主要科研机构已非常普及, 基本可以满足物相鉴定及简单晶体结构测定等基本需求; 我国已经建成投入使用的合肥国家同步辐射实验室和上海第三代同步辐射光源, 为复杂晶体结构的测定和物质精细结构的研究提供了分辨率更高的光源; 2008年已经立项的东莞散裂中子源项目的开工建设将为磁结构研究提供强大的硬件支持。可以说, 在物质结构测定硬件建设方面, 我国已达到或接近世界先进水平。然而, 在晶体结构测定软件技术发展方面, 我们通过对国际最大晶体学计算软件网站 CCP14 (<http://www.ccp14.ac.uk>) 的统计表明, 该网站收录了来

自全球 189 个研究组发布的晶体学相关计算程序，我国仅有中科院物理所范海福研究组和董成研究组的软件成果被收录，前者发布的是单晶结构解析相关软件(OASIS, VEC, DIMS, SAPI)，董成课题组发布的是粉末衍射结构解析相关软件 PowderX[24]，两者仅占全球晶体学分析软件总数的约 1%。也就是说，我国在晶体学分析软件发展方面非常薄弱，与世界整体晶体结构分析和研究水平有很大差距，更与当今我国所拥有的硬件设备（包括先进的上海光源和东莞散裂中子源）极不相适应。因而，从事和发展具有我国自主知识产权的晶体结构分析软件显得极其重要，这项研究有利于改变我国在在先进科学仪器和专用软件方面长期依靠进口的局面。

随着人工智能技术的发展，越来越多的复杂问题已经可以由人工智能算法加以解决，比如，2016 年基于深度神经网络人工智能算法的谷歌围棋程序 AlphaGo 连续击败世界多位围棋冠军[3]，证明了人工智能算法的强大威力。人工智能算法引入晶体学领域将有可能给解决复杂的晶体学问题带来新的思路。由此，本文拟采用人工智能算法中的专家系统[25-27]和 Q 学习算法[28-29]来提高 Rietveld 精修的自动化和智能化程度。利用专家系统算法，在精修策略库的基础上模拟人类专家的精修进程从而实现高水平的 Rietveld 精修自动化能力，Rietveld 精修时不需要人为干涉，精修结果发散和精修程序运行失败时，能自动回归之前的最佳值,而不是传统手动的 Rietveld 精修在出现发散时只能从头再来。同时利用强化学习算法（Q 学习算法）来搜索更优精修策略去丰富专家系统的精修策略库，Q 学习算法具有动态决策能力；有冒险机制，尝试不同的精修策略；探索最佳的精修策略，具有学习和记忆能力，对某类问题进行训练后，下次遇到类似问题可以更快的速度进行收敛。对于 Rietveld 精修自动化的程序的开发,选择目前被广泛使用的 Rietveld 精修程序 FullProf 作为 Rietveld 精修引擎，把程序开发的主要精力放在采用人工智能算法对现有 Rietveld 精修程序的自动化和智能化改进上，而非改造现有的 Rietveld 精修程序。这是本项目的特色和创新点。

此外，基于 Rietveld 方法的混合物定量相分析是材料衍射分析的重要内容[30]，但传统的 FullProf 定量相分析步骤多，操作繁琐。基于 AutoFP 的 Rietveld



精修自动化能力,继续开发基于 AutoFP 的自动化定量相分析程序,将大大简化采用 FullProf 进行定量相分析的操作复杂度。

随着材料基因组计划的提出,高通量的材料制备成为重点。随着高通量材料制备技术的推进,高通量的材料表征就应运而生[31-32],因粉末衍射是材料结构探测的经典方法,高通量的粉末衍射自然成为高通量材料表征的重要一环。高通量粉末衍射造成粉末衍射数据规模的急剧膨胀,传统的人工分析将面临极大地挑战,甚至已经无法应对如此大数据的汹涌而来。在这种情况下高通量的粉末衍射数据分析就变得非常急迫和重要。Rietveld 精修作为粉末衍射结构分析的重要方法,高通量的 Rietveld 精修自然而然就成为了研究的重要课题。因此,开发高通量的粉末衍射数据分析程序就变得非常迫切。因为 AutoFP 已经可以对单任务高度自动化地完成精修任务,这为高通量的精修创造了条件。基于自动化 Rietveld 精修程序 AutoFP 开发高通量的 Rietveld 精修系统,将为高通量 Rietveld 精修创造可能性,从而极大推动高通量材料研发技术和材料基因组工程的发展。

总之,该项研究是为了开发自动化和高通量的 Rietveld 精修程序[1-2],该项研究将人工智能算法引入 Rietveld 精修领域,以实现自动化和高通量的 Rietveld 精修,并把程序应用到新型材料探索和材料基因组工程的高通量衍射数据分析中,同时为一般粉末衍射精修研究也提供一个有力的计算工具。

## 1.2 国内外研究概况

### 1.2.1 国外研究概况

Rietveld 精修自动化的国外现状、水平及发展趋势。目前, Rietveld 精修程序使用比较广泛的主要是 DBWS, FullProf, GSAS, Rietan 等,这些程序的一个主要特点就是自动化水平较低。美国哥伦比亚大学的 Simon Billinge 教授出现在 Rietveld 精修程序的“三宗罪”:超级复杂费时费力(complex and labor intensive),糟糕的收敛策略(terrible convergence),不能快速处理多个精修[2,33]。在美国国家自然科学基金的资助下, Billinge 教授研究组开始开发 SrRietveld 精

修自动化程序[2]，该程序的文章于 2013.1 月正式发表，但从其文章中的描述来看，他们的把主要精力花在了原有程序的改写上面，他们把由 Fortran 代码编写的 FullProf 和 GSAS 程序改成由 Python 代码来编写；在文章中他们对 SrRietveld 程序的应用仅仅是对 NaCl 结构的一个自动化测试，可以说他们还没有实现真正的自动化。目前，除 Billinge 教授的研究组外，还没有其他研究组开展 Rietveld 精修自动化这类工作。本文使用人工智能算法来提高 Rietveld 精修的自动化水平，虽然 Rietveld 精修自动化这个想法并非首创(显然，这个想法是大多数使用 Rietveld 精修程序的科研人员的普遍需求)，但把人工智能算法应用到 Rietveld 精修的自动化领域属于本文研究的创新之处，目前还没有把人工智能算法应用于 Rietveld 精修自动化的文献报道。

### 1.2.2 国内研究概况

在庞大的国际晶体学软件库[34]里面少有国产软件，关于 Rietveld 精修自动化的探索国内也没看到相关方面的研究报道。一方面开发国产软件，增强国内晶体学软件研发能力势在必行，另一方面随着中国版材料基因组计划的推出[32]，面对高通量材料表征所产生的海量数据，开发高通量的分析软件将变得非常重要，而自动化的实现将成为实现高通量的基石。

## 1.3 论文的主要研究内容

本论文是以作者攻读硕士学位期间所承担课题的工作为基础，在第一章中阐述了课题研究的来源、目的、意义以及国内外研究的现状。第二章阐述了基于人工智能算法的 Rietveld 精修自动化程序 AutoFP 的程序结构，原理和实现。第三章阐述了 Rietveld 方法定量相分析的原理和基于 AutoFP 的自动化定量相分析的实现。第四章阐述了高通量精修的提出以及基于 AutoFP 和 MPI 并行化框架的高通量 Rietveld 精修系统。第五章阐述了对以上各个程序和系统的测试和结果讨论。最后，第六章对全文做了总结并做出了展望。

## 第二章 基于人工智能算法的自动化

### Rietveld 精修程序 AutoFP

#### 2.1 AutoFP 程序架构

##### 2.1.1 程序架构

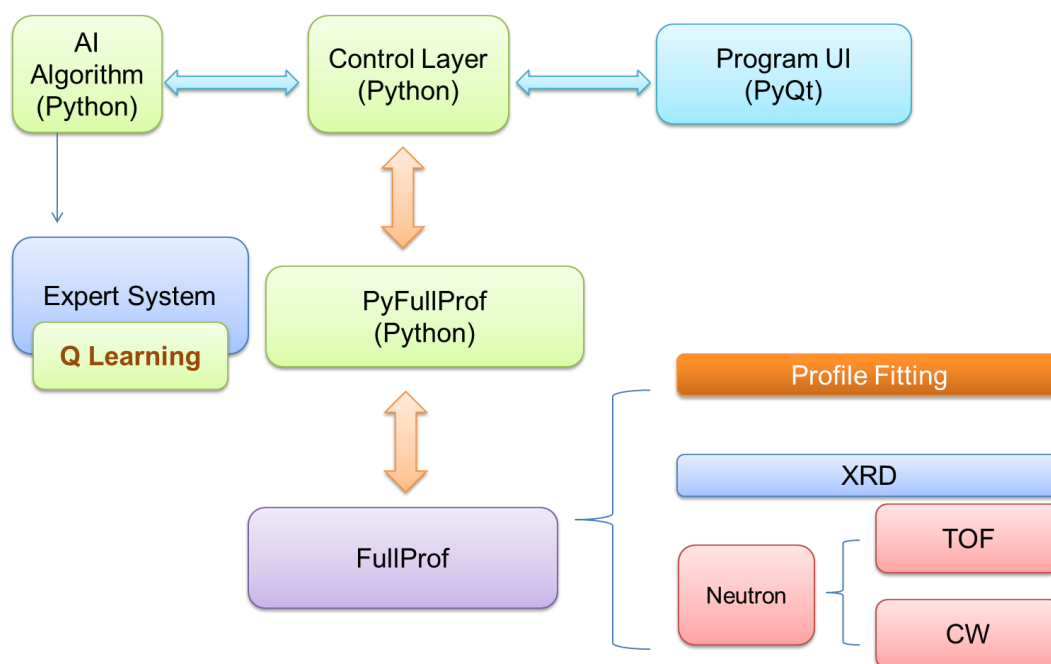


图 2-1 AutoFP 程序架构

要使 FullProf 的精修过程自动化，第一步必须能够正确地解析 FullProf 的输入输出文件，主要包括精修控制文件（pcr 格式文件），输出文件（out 格式文件，sum 格式文件等等），但由于 FullProf 是一个功能多复杂度高的程序，因此这些文件的结构也是异常复杂，从头编写程序解析这些文件不仅要耗费大量精力，而且正确性还需要反复验证，也耽误了主要的自动化的目的，寻求现有开源解析代码将是一个捷径。SrRietveld 就是这样一个工程[2]。SrRietveld 采用 Python 代码编写，BSD 许可协议，开源免费，提供了一个很好的 FullProf 接口 PyFullProf，可以正确地读写 FullProf 的控制文件和输出文件。因此，AutoFP 将把 PyFullProf 模块作为与 FullProf 的交互层。控制层的目的是控制 PyFullProf

按精修进程读写指定文件，并接受来自 AI 模块的控制命令，控制层也必须具备调用 FullProf 精修核心程序 fp2k 的能力，控制 fp2k 的启动，结束，并具有对 fp2k 的运行错误进行探测的功能，能将精修结果或错误代码反馈给 AI 模块。控制层起到了隔离精修过程和 AI 模块的作用，有了控制层 AI 模块可以抛开具体的问题，而只面对一个抽象的接口进行学习和优化，寻求最优值。AI 模块是自动精修的决策核心，负责策略调度，精修决策，寻找最优值，第一级专家系统模块，采用人工智能专家系统的架构，将精修策略做成策略库，根据不同的精修任务，调用不同的策略从而模拟人类专家完成精修过程，第二级策略搜索模块，采用机器学习的方式寻找最优策略为第一级专家系统策略库补充精修策略。AutoFP 这种模块化的设计简化了整个程序的复杂度，控制层将决策算法和具体精修引擎分离，极大地拓展了 AutoFP 的自由度，对于 AutoFP 以后的开发，可以选择性替换决策算法模块和精修引擎模块，比如可以替换 PyFullProf 为 PyGSAS 从而把精修引擎替换为了另一个 Rietveld 精修程序 GSAS，而算法模块无需改变。也可以替换算法模块为其他机器学习算法，而不影响精修引擎模块。

AutoFP 目前使用 FullProf 作为精修引擎，但保留了 GSAS 作为精修引擎的接口，方便未来拓展。决策算法部分，专家系统目前是主要算法，在专家系统基础上正积极发展第二级基于机器学习的策略搜索算法。

### 2.1.2 跨平台性

AutoFP 所依赖的 Python[35]，PyQt[36]，Matplotlib，Numpy 等开源 Python 库，以及 Rietveld 精修软件 FullProf 都是跨平台(Windows，Linux，Unix，Mac)的，所以 AutoFP 也就拥有良好的跨平台性。目前已经移植到 Windows，Linux 两大主流平台。Windows 平台和 Linux 平台 AutoFP 如图 2-3 和图 2-4 所示。其中 Linux 平台的移植为后续高通量 Rietveld 精修创造了条件。

命令行模式：`autofp -c n -a *.pcr`

`-c n` 运行 `n` 个 cycle，如果 `n=0`，自动决定运行 cycle，`-a` 自动选择精修参数

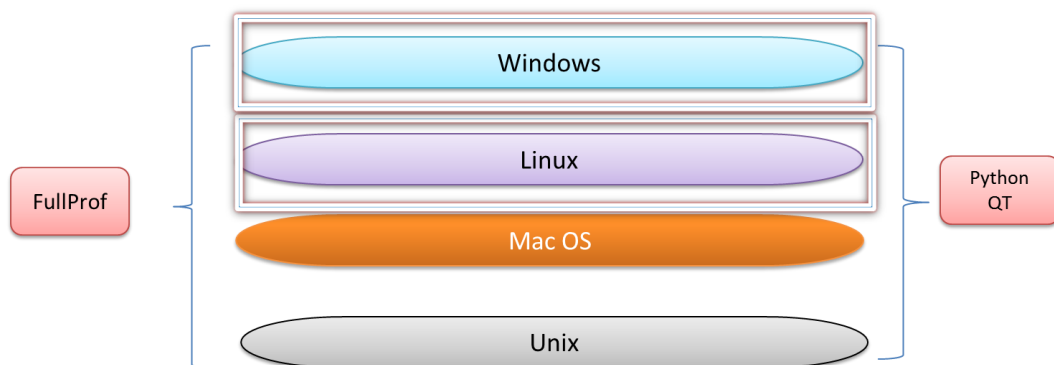


图 2 - 2 AutoFP 跨平台性

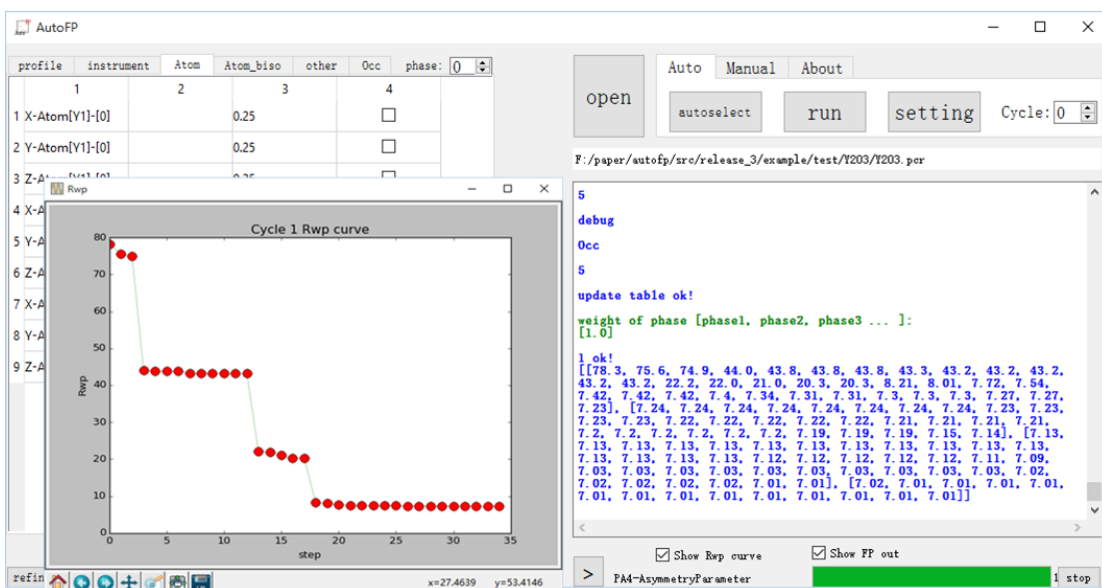


图 2 - 3 Windows 平台 AutoFP 界面

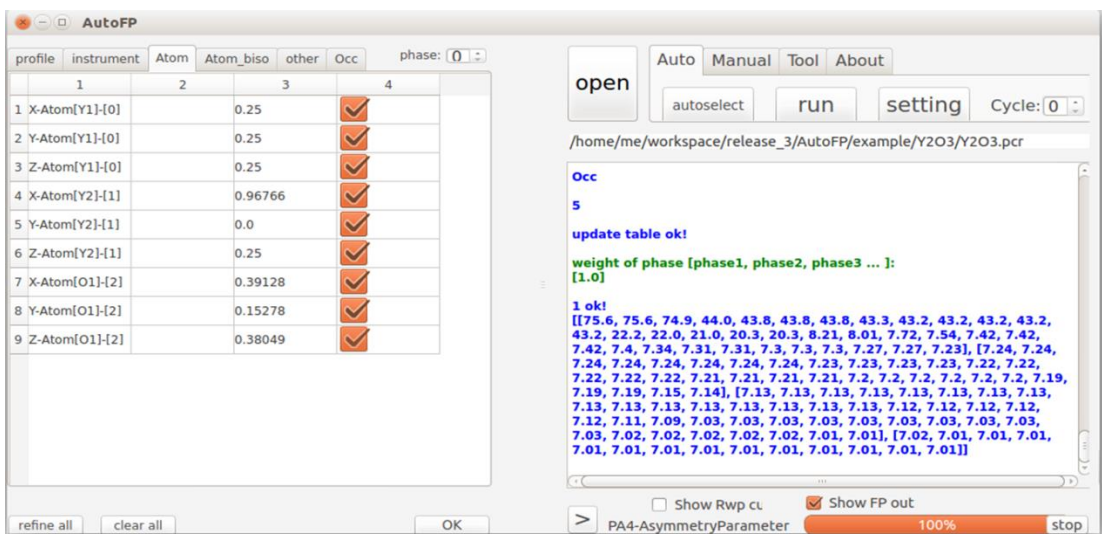


图 2 - 4 Linux 平台 AutoFP 界面

### 2.1.3 设计原则

AutoFP 采用面向对象的 Python 语言开发[1,35]。面向对象的设计使得 AutoFP 可以被轻松地管理和扩展。因为其跨平台和易用性, Python 语言在科学软件开发领域被广泛采用。NumPy (<http://www.numpy.org>) 和 Matplotlib (<http://matplotlib.org>) 是适用于数值分析和可视化被广受欢迎的 Python 软件包。在 AutoFP 开发过程中, Matplotlib 软件包被采用来做动态  $R_{wp}$  曲线图像显示。Numpy 包被用来做复杂的数值计算和矩阵分析。AutoFP 的用户界面 (UI) 是采用 PyQt 包[36] (<http://sourceforge.net/projects/pyqt>) 开发的。Qt 框架是被广泛采用的 UI 设计框架, 具有良好的跨平台特性。PyQt 是 Qt 框架的 Python 接口。Python 和 Qt 框架的这些特征可以帮助 AutoFP 具有良好的跨平台性。

AutoFP 的第一级控制算法采用专家系统算法[25,27]。专家系统算法是一种经典的人工智能算法, 被设计用来解决基于 IF-THEN 规则的知识推理的复杂问题。知识是基于条件和规则的。专家系统的推理引擎把规则应用于已知的状态从而推进系统进入新的状态。在 Rietveld 精修过程中, 人类专家的精修策略可以作为专家系统的知识库。此外, 推理引擎也可以拥有解释和调试能力。

AutoFP 的设计原则[1]是:

- 1) 模拟人类专家的决策能力去完成精修任务
- 2) 使精修过程高度自动化
- 3) 源代码开放, 可扩展

对于原则 1), 为了模拟人类专家去完成精修任务, AutoFP 采用专家系统算法被。在 AutoFP 里面, FullProf 精修软件被选为精修引擎。决策算法是基于专家系统算法的推理引擎。知识库是已经存在的精修经验和精修策略, 把这些经验和策略简化为 IF-THEN 规则, 这些规则构成了精修的知识库。当 AutoFP 开始一个精修任务时, 它的控制层将会按照知识库自动地安排所要精修的参数的顺序, 并监控 FullProf 的运行状态, 基于精修知识库处理精修过程中遇到的各

种精修问题，比如精修发散，FullProf 运行意外停止等等。从而，去高度自动化地完成复杂的 Rietveld 精修任务。

对于原则 2)，为了简化和自动化精修进程。AutoFP 被设计成界面友好的 UI 模式，尽量减低用户的输入和选择操作。为了降低学习精修新用户的学习难度。AutoFP 携带了一些参考精修知识库，用户无需自己构建知识库便可以完成很多精修任务。当然，AutoFP 保留了用户自定义精修策略的接口，高级用户可以写自己的策略，从而调动 AutoFP 去完成更复杂的精修任务，用户同时也可以分享自己的好的精修策略。控制层保存并监视 FullProf 运行过程中的各种中间状态，如果 FullProf 精修过程中遇到发散或者 FullProf 运行错误，控制层将会按照专家系统精修策略知识库自动地从保存的历史步骤中恢复最好的状态，从而恢复精修进程。通过这种方式，AutoFP 将会试图去最小化精修任务的  $R_{wp}$  值。

对于原则 3)，AutoFP 将作为 GPLv3 许可下的开源项目，有兴趣的用户可以获取源代码，修改代码来满足自己的需求，也可以贡献自己的代码促进 AutoFP 的发展。高级用户可以开发自己的精修策略，或者改变 AutoFP 的决策算法去提高精修质量。

#### 2.1.4 使用说明

AutoFP 简单易用，用户可以按照以下步骤轻松使用。

1) 加载程序 AutoFP

2) 选择“Auto”标签页

3) 点击“open”按钮打开 pcr 格式精修控制文件，或拖拽 pcr 格式文件到界面上自动打开。点击“autoselect”按钮去自动选择精修参数，或者在左侧参数界面手动选择精修参数选择模式，“cycle=0”表示自动决定循环次数，“cycle=n”表示程序将运行 n 个循环

此外，使用 AutoFP 需要注意的地方：

### 1) 请制作正确的精修控制文件

使用 FullProf 制作正确的精修控制文件, AutoFP 运行需要精修控制文件 (\*.pcr) 和数据文件 (\*.dat)。请确保 pcr 文件和 dat 文件拥有相同的名字, 比如 Y2O3.pcr, Y2O3.dat。请将这两个文件放置于一个新建的空文件夹下, 以方便程序控制。由于 Qt 库中文支持的因素, 文件路径名中文支持还有问题, 因此最好使用英文路径名。

2) AutoFP 随身携带了一些测试例子可以在 AutoFP 网站 <http://pmedia.shu.edu.cn/autofp> 下载或者在 AutoFP 安装目录下获取, 比如 C:\Program Files(x86)\AutoFP\AutoFP\example.zip。

更多资源和视频教程可以访问 AutoFP 网站 <http://pmedia.shu.edu.cn/autofp> 获取。

## 2.1.5 附加工具集

AutoFP 除了提供自动 Rietveld 精修功能外, 还提供了一些附加工具, 以方便用户使用。比如自动 prf 文件到 origin 转换的功能和自动由 cif 生成 pcr 文件的功能

### 1) prf 文件到 origin 文件自动转换工具 prf2origin

FullProf 输出的精修结果图像文件是由 FullProf 定义的 prf 格式, 此格式只能用 Fullprof 软件套装的 Winplotr 打开观看。但实际科研工作中往往采用 origin 软件来做图像展示, 手工从 prf 提取数据到 origin 软件里进行调整画图是一个繁琐低效的过程。为了方便用户展示精修结果, 本人开发了一个自动化的 prf 到 origin 的转换工具, 可以极大便利科研工作的需要。

转换过程由 Python 代码将 prf 文件转化为 origin 能识别的数据文件, 再调用 origin C 函数接口用 origin 对数据文件绘图, 绘图借助已制作的 origin 图像模板自动生成, 该模板可由用户自定义改变。

### 2) cif 自动生成 pcr 文件工具 makepcr



cif 格式文件是国际晶体学会通用的晶体结构文件，而 FullProf 精修采用的是自己定义的精修控制文件 pcr 格式。因此欲采用 FullProf 精修，通常有一个 cif 文件到 pcr 文件的转换过程，之前只能采用 edpocr 这个 FullProf 工具套装提供的界面方式进行转换，不但效率低，而且考虑到后面的高通量精修的要求，这种方式很不方便。基于此，本人采用 Python 正则表达式工具，开发了自动化的 cif 到 pcr 的转换工具。

由正则表达式功能解析 cif，从 cif 晶体结构文件取出晶体结构信息（空间群，晶胞参数，原子坐标等），然后用正则表达式匹配已经制作的 pcr 精修控制文件模板，用提取出的晶体结构替换其中的晶体结构部分，从而完成自动由 cif 晶体结构文件制作 pcr 精修控制文件的过程。

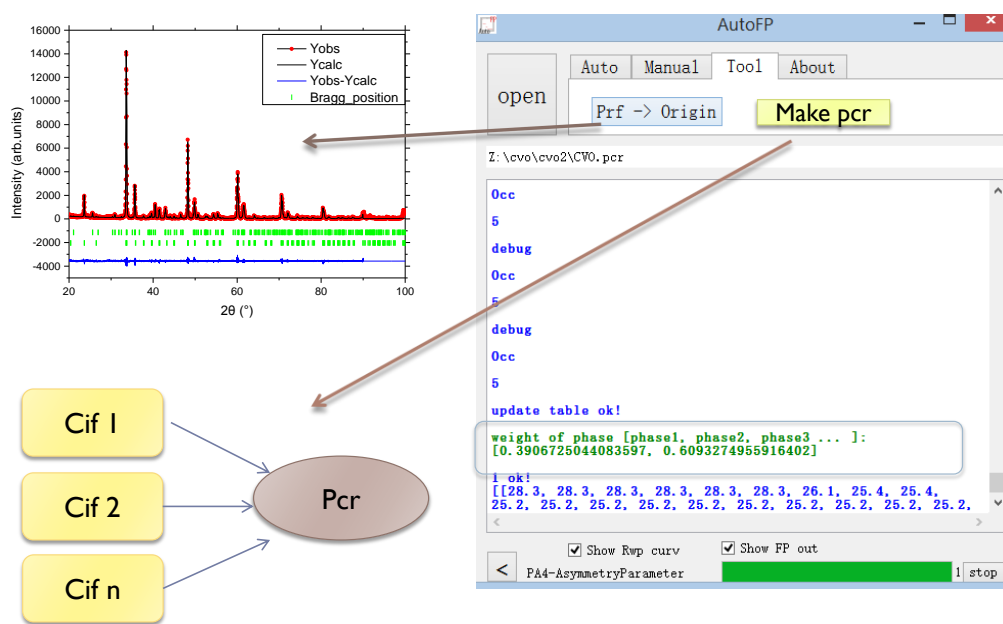


图 2-5 AutoFP 里面的附加工具集 prf2origin 和 makepcr

## 2.2 专家系统

### 2.2.1 专家系统的原理

专家系统[25-27]是一种人工智能程序系统。专家系统内部包含有大量的某个领域专家级的知识与经验。它利用人工智能技术，根据某领域专家提供的知识和经验，进行推理和判断，模拟人类专家的决策过程，以便解决那些需要人

类专家处理的复杂问题，简而言之，专家系统是一种模拟人类专家来解决特定领域问题的计算机程序系统[37,38]。

专家系统通常由 6 个部分组成，推理机、知识库、人机接口、解释器、综合数据库和知识获取[38]。其中知识库与推理机相互分离是专家系统的一大特色。根据所要解决的问题类型和规模的不同，专家系统的整体结构有所差异。

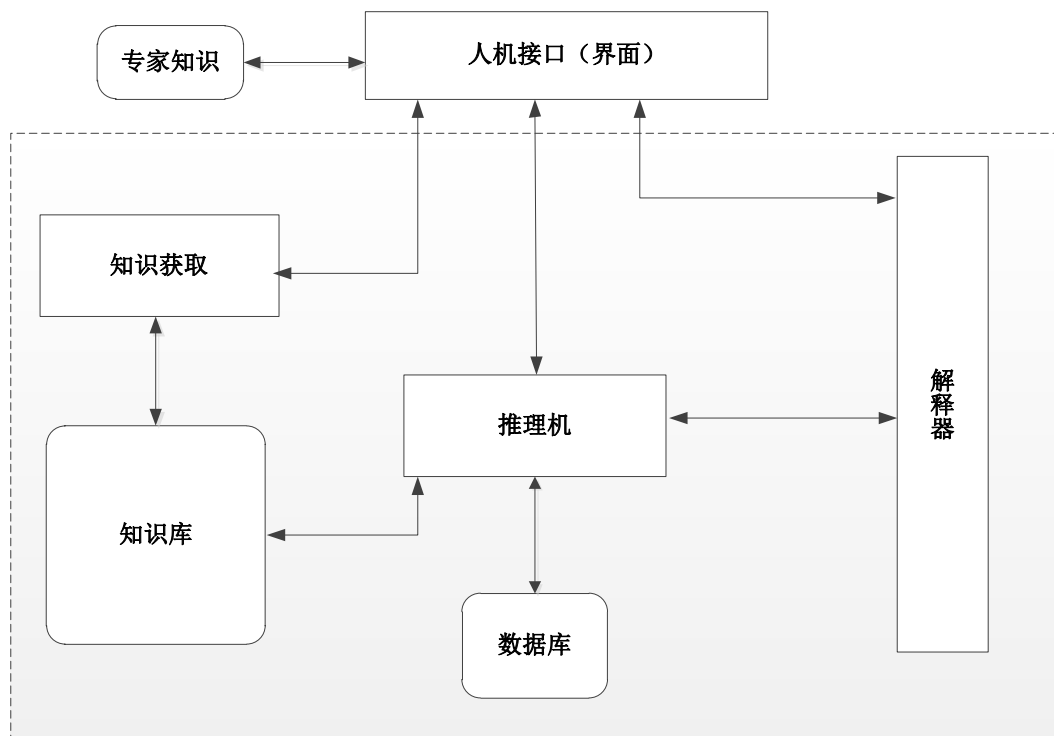


图 2 - 6 专家系统结构

为了使专家的领域知识能够在计算机中加以利用，必须将知识采用合适的形式加以表示。产生式规则、语义网络、逻辑模式等等都是目前常用的知识表示方式。其中基于规则的产生式系统是目前知识表示中最为常用的方法。这一方法中知识库由用“IF-THEN”形式表达的大量知识规则构成[25]。

推理机，又叫规则解释器，其任务是运用控制策略去寻找可以应用的规则。推理机针对当前状态或信息，反复匹配存储于知识库中的规则，获得新的结论，使系统进入新的状态，最终求解问题[25]。

知识库是用来存放领域专家提供的专业知识。利用知识库中的专家知识，专家系统进行知识推理求解问题，从而模拟专家的思维方式。所以知识库的好坏对专家系统的整体水平具有决定性作用。专家系统具有知识库与推理机是相互分离的优点，因此用户可以通过提升知识库来提高专家系统的水平，而无需改变推理机。知识库建设除了通过收集专家知识外，还可以利用机器学习的方式自主寻找更好的知识。

人机接口（界面）负责用户和系统进行交流。通过人机接口，用户输入问题，系统输出推理结果和相关的解释过程等。

综合数据库专门设计用于存储系统推理过程中所需的原始数据、产生的中间结果，以及最终的结论，往往用来作为暂存区[25]。

解释器能够根据用户的提问输入和推理系统的运算结果，给出求解过程和结论说明。从而使专家系统推理具有可回溯性，也更有人情味[25]。

知识获取负责提升专家系统知识库的知识水平，因而成为专家系统建设的关键问题。传统的专家系统通过手动输入专家的知识来组件知识库，知识获取成为瓶颈问题。然而利用机器学习算法自动寻找新的知识可以突破这一瓶颈。本文正是利用机器学习算法寻找新知识来提升专家系统水平。

### 2.2.2 专家系统在 AutoFP 里的实现

Rietveld 精修可以看做一个决策树过程，Rietveld 精修难点集中在初学者精修策略不当，精修频繁遇到发散崩溃等问题，且不容易恢复。对于这些困难利用一些积累的经验是可以解决的。借助专家系统的概念，将这些经验编入程序[27]形成知识库将极大简化精修过程。基于专家系统的自动化精修程序框架如图 2-7 所示：

精修知识库由精修策略构成。关于策略的定义 Rietveld 精修往往包含许多参数。在 AutoFP 里面一种精修参数顺序的安排，错误处理动作，和追求的目标函数的综合，我们称之为一种策略。很显然 Rietveld 精修作为典型的多参拟合过程，不同的精修策略将会产生非常不同的精修结果。一些策略可能是非常完美的，但更多的将会是很糟糕的。AutoFP 被设计成采用一种指定的精修策略去

自动化地完成精修任务。这个策略可以被用户根据他们的需要指定或者修改。为了方便新用户，AutoFP 为 X 射线衍射和中子衍射（CW 模式，TOF 模式）携带了一些基于文章[39-40]建议的默认精修策略。

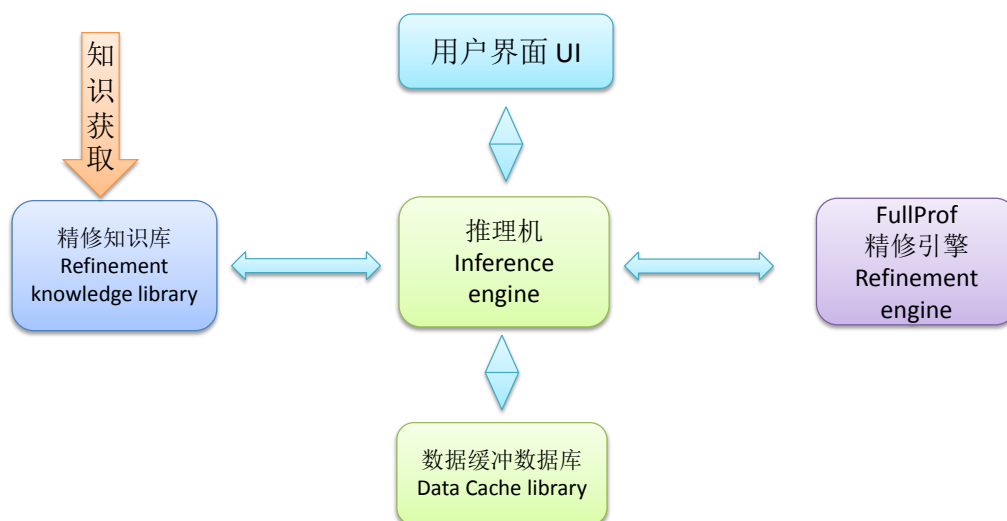


图 2-7 基于专家系统的自动化 Rietveld 精修程序结构

AutoFP 默认的精修策略将会按照先精修峰形参数，再精修晶体结构参数，最后精修择优取向的顺序，如果出错则自动返回，追求 $R_{wp}$ 最小化的要求去完成精修任务。具体的精修参数勾选顺序（X 射线衍射模式）如下所示：

[ 2 $\theta$  zero correction, unit cell parameters, sample displacement, simple background correction, scale factor, peak width, peak shape function, peak asymmetry, complex background correction, atomic positions, atomic displacement parameters, occupancy parameters, preferred orientation ] .

尽管 AutoFP 提供的这些默认策略可以解决一些通用的 Rietveld 精修任务。但显然这个默认的策略不可能满足所有用户的需求。所以用户必要时写自己的精修策略将会是非常有益的。基于 AutoFP 的开放性设计，修改 AutoFP 的默认精修策略或者书写自定义的精修策略将会是非常容易的事。

表 2-1 基于专家系统的自动化 Rietveld 精修算法流程

1. Read the PCR file and decide the type of Rietveld refinement (XRD, Neturon-CW, Neturon-TOF)
2. Select the refinement parameters and configure the automatic options
3. Start automatic Rietveld refinement

Loop begins:

Read the refinement strategy from knowledge library, save the PCR file, the parameters of Rietveld refinement consist of three main classes: profile parameters, structural parameters and preferred orientation.

Sub-loop begins:

Use expert system to Check next parameter and run FullProf

a. Profile parameters

peak shape: peak shape function, peak width as a function of  $2\theta$ , peak width as a function of hkl (anisotropic line broadening), peak asymmetry as a function of  $2\theta$  (axial divergence), peak position: zero point error, sample displacement, unit cell parameters background

b. Structural parameters

scale, atomic positions, atomic displacement factors, atomic occupancy parameters Occ

c. Sample parameter preferred orientation

if:  $R_{wp}$  ascends or *FullProf* meets error

then: backup the PCR file

else: save the new PCR file

End of the sub-loop

End the Loop (when  $R_{wp}$  convergence or the specified number of cycles)

4. End the program

### 2.2.3 精修策略

正则表达式(regular expression)RE 是一种被广泛使用的字符串搜索工具。RE 描述了一种字符串匹配模式[41]，该匹配模式由一种字符串序列构成，序列由普通字符和被称为“元字符”的特殊字符组成，模式描述了文本搜索时要匹配到的字符串的结构规则[41]。在字符串搜索时，RE 充当一个模板，在所搜索的文本中寻找符合其匹配模式的字符串。RE 可以用来检查一个文本中是否包含有某种模式的子串，还可以对匹配到的子串进行替换或提取等操作。大多数高级编程语言（如 C++，Java，Python 等）都能很好地支持 RE。构成 RE 的组件主要有单个字符或字符集合、字符范围或字符位置等等[42]。构造 RE 的过程与构造数学表达式的过程比较类似，除了可以利用上述组件组合来生成一个小的正则表达式，还可以利用运算符将小的表达式结合在一起制作更大的表达式。

AutoFP 执行精修策略所指定的参数顺序使用的是正则表达式匹配。AutoFP 使用参数名前缀去匹配一批参数，从而按策略顺序精修。例如，“X-Atom” 匹配 “X-Atom[1], X-Atom[2], ...”；“BACK” 匹配 “BACK[0], BACK[1], ...”；而 “BACK[0]” 只匹配 “BACK[0]”。

为了方便用户自定义策略，以及调用由机器学习算法所生成的策略，AutoFP 定义了一种策略文件，只要修改策略文件，就可以改变 AutoFP 的精修策略。精修策略文件构成了 AutoFP 的专家系统知识库。策略文件分为 X 射线衍射，中子衍射 CW 模式，中子衍射 TOF 模式，峰形拟合，定量相分析等多种任务模式。策略文件主要包含了一下几个部分，任务类型说明，精修参数勾选顺序，目标函数，FullProf 崩溃发散等异常时的处理动作。任务类型支持 XRD, Neutron\_CW, Neutron\_TOF，峰形拟合等多种任务类型。目标函数定义了用户所希望追求的最优化目标，比如  $R_{wp}$  最小化， $R_p$  最小化，或者其他的一些组合。精修参数顺序，用户可以根据需求自定义顺序，或者由机器学习算法搜索好的策略顺序保存使用。异常处理机制控制 AutoFP 在 FullProf 运行出现异常后的动作，比如撤销操作，修改参数等。

表 2 - 2 AutoFP 里 X 射线衍射所有可以精修的参数示例

[Zero-Pattern[0], Sycos-Pattern[0], Sysin-Pattern[0], Lambda-Pattern[0],  
 BACK[0]-Background, BACK[1]-Background, BACK[2]-Background,  
 BACK[3]-Background, BACK[4]-Background, BACK[5]-Background,  
 X-Atom[0], Y-Atom[0], Z-Atom[0], Occ-Atom[0],  
 Biso-AtomicDisplacementFactor, B11-AtomicDisplacementFactor,  
 B22-AtomicDisplacementFactor, B33-AtomicDisplacementFactor,  
 B12-AtomicDisplacementFactor, B13-AtomicDisplacementFactor,  
 B23-AtomicDisplacementFactor, X-Atom[1], Y-Atom[1], Z-Atom[1],  
 Occ-Atom[1], Biso-AtomicDisplacementFactor,  
 B11-AtomicDisplacementFactor, B22-AtomicDisplacementFactor,  
 B33-AtomicDisplacementFactor, B12-AtomicDisplacementFactor,  
 B13-AtomicDisplacementFactor, B23-AtomicDisplacementFactor,  
 X-Atom[2], Y-Atom[2], Z-Atom[2], Occ-Atom[2],  
 Biso-AtomicDisplacementFactor, B11-AtomicDisplacementFactor,  
 B22-AtomicDisplacementFactor, B33-AtomicDisplacementFactor,  
 B12-AtomicDisplacementFactor, B13-AtomicDisplacementFactor,  
 B23-AtomicDisplacementFactor, Scale-Contribution[0], Shape1-Profile,  
 Bov-Contribution[0], Str1-StrainParameter, Str2-StrainParameter,  
 Str3-StrainParameter, U-Profile, V-Profile, W-Profile, X-Profile, Y-Profile,  
 GausSiz-Profile, LorSiz-Profile, a-Phase[0], b-Phase[0], c-Phase[0],  
 alpha-Phase[0], beta-Phase[0], gamma-Phase[0],  
 Pref1-PreferOrient, Pref2-PreferOrient,  
 PA1-AsymmetryParameter, PA2-AsymmetryParameter,  
 PA3-AsymmetryParameter, PA4-AsymmetryParameter,  
 S\_L-AsymmetryParameter, D\_L-AsymmetryParameter]

```

'''strategy setting file, note: this is an important file. You
should be careful when modifying the file.
Please keep the format.
The words is case sensitive.
You can modify parameters group order in 'param_order' and
parameters group in 'param_group'.'''
strategy={
  "xray":{
    # task type
    "type":"xray",
    # param group format:
    'group_name':[ group_member1,group_member2,...]
    "param_group":{
      'scale': ["Scale"],
      'zero': ["Transparency","Zero"],
      'simple background': ["BACK[0]","BACK[1]"],
      'cell a,b,c': ["a-Pha","b-P","c-P"],
      'W': ["W-Pr"],
      'complex background': ["BACK"],
      "V,U": ["V-Pr", "U-Pr"],
      "Asym": ["PA"],
      'Y,X': ["Y-Pr","X-Pr"],
      'Atom x y z': ["X-Atom","Y-Atom","Z-Atom"],
      'Pref,Bov': ["Pref","Bov"],
      'Biso-Atom': ["Biso-Atom"],
      'GausS,1G': ["GausS","1G"],
      'Occ-Atom': ["Occ-Atom"],
      'Temperture factors': ["B1","B2","B3"],
      'S_L,D_L': ["PA", "S_L","D_L",
                  #"Sysin","Displacement",
                  #"LorSiz","SZ",
                  #"Sycos","Transparency",
                  #"Str",
                  ],
      "manual background": ["BCK"]
    },
    # param order format: [group_name1,group_name2,...]
    'param_order': [
      "scale",
      "zero",
      "simple background" ,
      "cell a,b,c",
      "W",
      "complex background" ,
      "V,U",
      "Asym",
    ]
  }
}

```

图 2 - 8 AutoFP 所使用的 Rietveld 策略文件



## 2.3 基于人工智能算法的 Rietveld 精修策略搜索

### 2.3.1 Q 学习算法搜索原理

Rietveld 精修可以被看作是一个马尔可夫决策过程(MDP)。马尔可夫性过程是这样一种随机过程，在知道目前状态的情况下，其未来演变不依赖于之前的历史[26]。马尔可夫决策过程是指在具有马尔可夫性的随机动态系统中，决策者周期地或连续地观察状态，并序列地做出决策[26,43,44]。每一步，决策者根据目前的状态，从可用的动作集中概率性地选用一个动作做出决策，系统进入下一步的状态（未来状态），其状态间转移概率是马尔可夫性的。在进入新的状态后，决策者继续作出新的决策，以此类推反复此过程。Q 学习算法（Q-learning）[28]是由 Watkins 在 1989 提出的一种用于解决 MDP 问题的强化学习算法。Q 学习算法的优点在于，它不需要依赖外部模型，这个特点对自动 Rietveld 精修问题非常有利。Q 学习的基本思想是，将状态和动作看作一个二维表的行和列，这样每个状态和动作都能形成一个状态-动作对，对应这张二维表中的一个元素，即 Q 值，这张二维表就是一张 Q 值表。决策者通过优化 Q 值表以达到最优决策的目的。决策时，决策者根据当前的状态选取 Q 值最高的状态-动作对进行下一步的动作。在转换到下一个状态时，决策者会收到相应的报酬，并据此更新相应的状态-动作对的 Q 值。Q 值的更新规则是：

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha_t(s_t, a_t) \times [r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$$

公式 2-1

其中：

- $Q(s_t, a_t)$  是当前的 Q 值
- $\alpha_t(s_t, a_t)$  是当前时刻的学习率， $\alpha_t$  决定了决策者的对报酬的敏感程度。

$$0 < \alpha_t < 1$$

- $r_{t+1}$  表示决策者在  $t+1$  时刻获得的报酬
- $\max_a Q(s_{t+1}, a)$  表示了在下—状态  $s_{t+1}$  能获得的最大 Q 值
- $\gamma$  是折扣率，表现了决策者对将来报酬的重视程度， $0 < \gamma < 1$ 。如果  $\gamma = 0$ ，则说明决策者不关心将来的报酬；如果  $\gamma$  接近 1，则说明决策者看重

长远的收益。

### 2.3.2 Q 学习算法精修策略搜索的实现

为了让 Q 学习算法能应用于 AutoFP 自动 Rietveld 精修，需要对 Q 学习算法进行如下调整

(1). 在选择动作的时候需要有冒险机制，每次从可用行动集中选择动作时，不单单选择一个拥有最大 Q 值的动作，而是根据用户的要求，选择一定数量的随机动作和多个拥有最大 Q 值的动作。由于所有的 Q 值初始值相同，一旦某个动作的 Q 值得到有利的更新，那么无疑每次在这个状态下都会去选择这个动作，其他动作就无法被尝试，它们的 Q 值也就无法被更新。因此，除了参考 Q 值表，自动化程序在选择动作时需要一定的冒险；

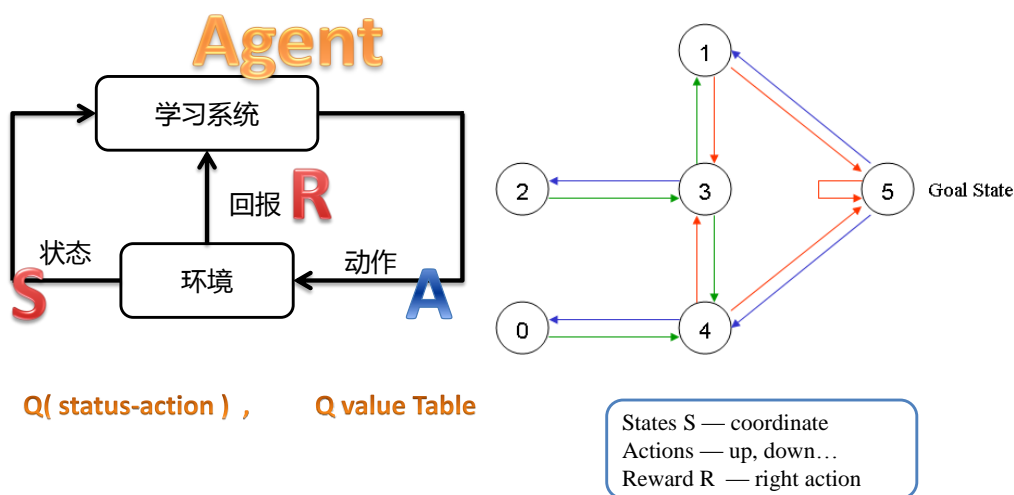
(2). AutoFP 已经实现了自动化地控制 FullProf 进行精修，只需完成 Q 学习算法与 AutoFP 控制层的对接即可实现基于 Q 学习算法的精修策略搜索。具体对接要完成以下几点 1) Q 学习算法的状态和动作要和控制层的精修状态和精修控制命令对接；2) Q 学习算法的奖惩回报要和控制层返回的精修质量和错误状态对接；3) Q 学习算法所搜索到的最优策略要生成 AutoFP 规定的策略文件以供专家系统算法所调取。

(3). 由于使用 FullProf 进行精修时很容易出现结果发散，所以需要 Q 学习算法进行策略尝试时需具备在结果发散后恢复的能力。一个简单的方法就是保存运行过程中所经历的状态，以确保随时能恢复到一个可用的状态。这点控制层已经可以实现，因此只要完成和控制层这一功能的对接即可。

综合上面的要求，Q 学习算法通过与 AutoFP 控制层交互，从控制层获取参数状态，精修质量，错误信息，通过控制层完成参数的勾选和去勾选，并从控制层获取精修结果作为奖惩依据。最终生成可供专家系统调用的精修策略文件。其学习过程如表 2-3 所示

SimpleAI 是著名的 Python 人工智能开源算法库，其提供了 Q 学习算法的完整代码，借助对 SimpleAI 的 Q 学习算法的改进，可以快速完成基于 Q 学习算法的 Rietveld 精修策略搜索程序。Q 学习算法通过反复尝试学习而获取的精修策略，将存储为策略文件而存储到第一级专家系统精修策略库，从而不断提高专家系统性能，完成第二级人工智能 Q 学习算法和第一级专家系统算法的无缝结合。通过 Q 学习算法搜索，AutoFP 有望超越人类专家能力。

### Q-Learning Algorithm



Q value update:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha (r_{t+1} + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))$$

图 2 - 9 Q 学习算法结构

表 2-3 Q 学习算法精修策略搜索算法流程

1. 获取初始状态  $S_{init}$ 。方法是固定所有参数后运行 FullProf，这样就获得了当前结构的  $R_{wp}$ 。
2. 将初始状态  $S_{init}$  存入状态集合  $S$  中。集合  $S$  是一个有序集合：每个插入集合  $S$  的状态都会被根据其  $R_{wp}$  值自动排序，以保证出现在集合  $S$  的第一个位置的状态元素拥有最小的  $R_{wp}$ 。
3. 循环执行以下步骤，直到满足一下三个退出条件：
  - a)  $R_{wp}$  达到预设的期望值
  - b) 循环次数达到最大值
  - c) 收到外部的停止指令循环体内的程序：
  - 1.1 从状态集合  $S$  中取出第一个状态（也就是  $R_{wp}$  最小的状态）作为当前状态
  - 1.2 获取适合当前状态的可用行动集  $A_{all}$
  - 1.3 从行动集  $A_{all}$  中随机取出若干个动作，放在执行动作集  $A_{exec}$  中
  - 1.4 从行动集  $A_{all}$  中取出拥有最大  $Q$  值的若干个行动，同样放在执行动作集  $A_{exec}$  中
  - 1.5 从  $A_{exec}$  依次取出动作，并循环执行以下动作
    - 1.5.1 将取出的动作作用到当前的状态上
    - 1.5.2 运行 FullProf，获取  $R_{wp}$
    - 1.5.3 如果 FullProf 的运行结果不收敛，则对这个动作进行“惩罚”，返回到上一步；如果获得有效结果，那么根据前后  $R_{wp}$  的变化更新当前状态和动作对的  $Q$  值。此次运行得到的  $R_{wp}$  和记录到的最小  $R_{wp}$  进行比较，如果此次运行得到的  $R_{wp}$  比最小的记录还要小，则保存当前的工作空间和  $R_{wp}$ 。

## 第三章 基于 AutoFP 的自动化定量相分析

### 3.1 FullProf 定量相分析

#### 3.1.1 Rietveld 定量相分析原理

粉末衍射定量物相分析是衍射分析的重要组成部分。1987 年 Hill 和 Howard 在研究 Rietveld 比例因子与相对质量分数的关系时指出[30], 从全谱拟合分析上来说, 每一相的质量分数与其 Rietveld 精修的比例因子、晶胞质量和晶胞体积是成比例的。如果混合样品中每一相均结晶良好, 并且样品中所含的相可以确定时, 混合样品中第  $i$  相的质量分数可以表示为:

$$w_i = \frac{S_i(ZMV)_i}{\sum_j S_j(ZMV)_j}$$

公式 3-1

式中,  $S$ 、 $Z$ 、 $M$ 、 $V$  分别表示 Rietveld 比例因子、晶胞的分子式单位、单位晶胞质量和晶胞体积。公式 3-1 中的  $w_i$  满足  $\sum w_i = 1$ 。但公式 3-1 没有考虑样品中各个物相的粒度与微吸收效应的关系。根据 Brindly 在 1945 年提出的关于粒度与相对吸收的关系式[30], Taylor 在 1991 年对公式 3-1 提出了关于吸收对比效应的修正式:

$$w_i = \frac{S_i(ZMV)_i/t_i}{\sum_j S_j(ZMV)_j/t_j}$$

公式 3-2

式中  $t_i$  是样品中第  $i$  相的吸收因子, 表达式为:

$$t_i = \frac{1}{A_i} \int_0^{A_i} e^{-(\mu_i - \bar{\mu})x} dA_i$$

公式 3-3

式中  $A_i$  是  $i$  相颗粒的体积,  $\bar{\mu}$  为线性吸收系数的中数。

### 3.1.2 FullProf 定量相分析功能

公式 3-2 与 FullProf 参数的对应是,  $S_i$  对应 Scale 参数,  $(ZM)_i/t_i$  对应 ATZ 参数,  $V_i$  是晶胞体积, 可由 FullProf 输出信息给出。因此 FullProf 中第  $i$  相的质量分数可以表示为

$$w_i = \frac{Scale_i(ATZ)_i V_i}{\sum_j Scale_j(ATZ)_j V_j}$$

公式 3-4

$Scale_i$  是第  $i$  相的 Scale 因子

多相精修 pcr 控制文件的制作。首先要准备所有相的 cif 晶体结构文件, 用 edpcr 把每个 cif 文件转为 pcr 文件, 然后合并所有 pcr 文件的 phase 部分内容到一个 pcr, 并修改这个 pcr 文件的 Nph 参数为相的数目。这个过程手动可以完成, 但为了方便用户, AutoFP makepcr 工具提供了自动实现了此过程, 可以将多个 cif 文件一键生成用于定量相分析的 pcr 多相文件。

FullProf 定量相分析, 随着相的增加, 需要精修的参数越来越多, 手动精修将变得越来越低效, 因此自动化这个过程并自动化计算相含量将变得非常迫切。AutoFP 是一个让 FullProf 精修过程实现自动化运行的工具, 在 AutoFP 的基础上稍加改动, 便可以实现这个目的。

## 3.2 AutoFP 自动化定量相分析

### 3.2.1 基于 AutoFP 自动化定量相分析的实现

AutoFP 的自动化定量相分析功能在 AutoFP 的 phase 模块实现, 其实现思路如下。FullProf 程序可以运行定量相分析任务, 为了是 AutoFP 能支持定量相分析, 需要对 AutoFP 内部数据结构进行拓展。具体实施就是将 pcr 文件中所有的参数在读入 AutoFP 时打上一个标志, 有该标志可以识别该参数属于第几个相。然后让 AutoFP 按照自动化精修的过程运行完毕, 最后读出每个相的 Scale, ATZ 参数, 并从 out 文件读入每个相的晶胞体积  $V$  参数, 根据公式 3-4, 由 AutoFP

的 phase 模块计算每一个相的质量分数，并在任务结束后将计算结果显示在 UI 界面上。

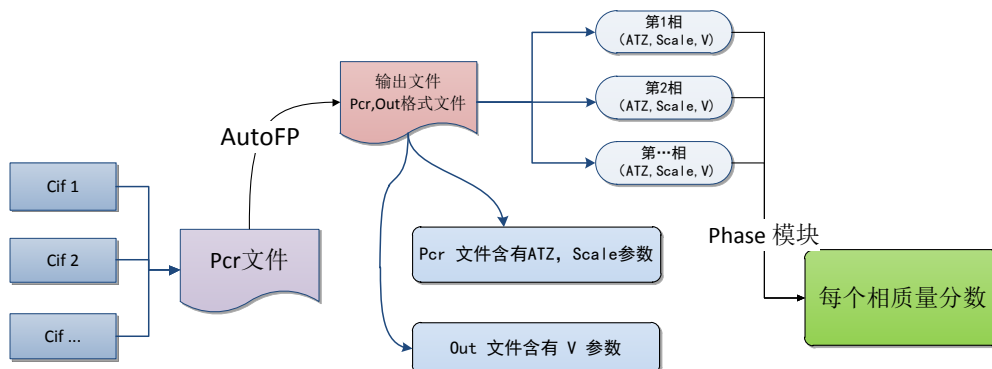


图 3 - 1 基于 AutoFP 的自动化 Rietveld 定量相分析流程

### 3.2.2 AutoFP 自动化定量相分析的使用

基于 AutoFP 的定量相分析使用步骤如图 3 - 1 所示，具体步骤如下

- 1) 按照 3.2.1 节中所述制作适用于多相定量相分析的 pcr 精修控制文件；
- 2) 按照 AutoFP 使用要求导入 pcr 文件，选择需要精修的参数，配置自动精修参数，然后自动精修；
- 3) 自动精修完毕，如图 3 - 2 界面将显示每一个相的质量分数含量，用户需要根据精修总体质量，决定是否接纳所计算出的质量分数

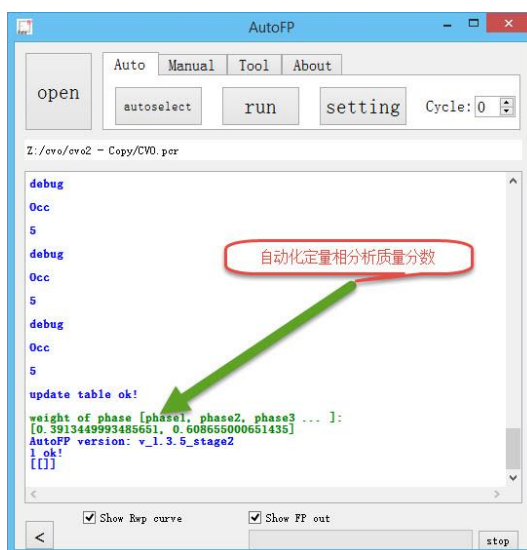


图 3 - 2 自动化定量相分析界面

## 第四章 基于 AutoFP 的高通量 Rietveld

### 精修系统

#### 4.1 高通量 Rietveld 精修的提出

材料基因组计划（即 Materials Genome Initiative），简称 MGI。2011 年 6 月 24 日，美国总统奥巴马在主题为“先进制造业伙伴关系”的演讲中提出了材料基因组计划，奥巴马明确地指出材料基因组计划的总目标是“将先进材料的发现、开发、制造和使用的速度提高一倍”[32]。白宫科技政策办公室当年发表的白皮书《具有全球竞争力的材料基因组计划》中阐述了材料创新基础设施的三个平台：计算工具平台、实验工具平台和数字化数据平台[32]。MGI 一经提出，便席卷全球，在全球材料研究领域产生了广泛影响。

材料基因组计划 MGI 是在信息技术革命后，美国充分认识到材料革新的重要作用，以谋求美国制造业复兴的战略背景下提出来的。自上世纪八十年代起，技术的革新和经济的发展已经越来越依赖于新材料的研发和应用[31]。目前，新材料从发现到最终应用一般需要一二十年的时间。显然，这一速度严重滞后于市场快速变化的需求。面对制造业的激烈变革和经济的快速发展，材料科学家和工程师必须缩短这一研发周期，以期来解决来自未来的挑战。然而，当前的新材料研发主要依据研究人员的科学直觉和大量炒菜式的重复性的尝试法实验，这种方法不仅效率低下，而且缺乏对材料的系统性研究[31]。此外，材料的发现，研究，应用，改良所涉及的各个团队之间彼此独立，缺乏合作和数据共享。而材料基因组计划提出的目的就是为了变革材料研发的方式，提升材料研发的效率。其中，高通量材料制备表征分析和高通量材料计算模拟预测成为了材料基因组工程的重要内容。

经过近些年的不断发展，高通量材料制备技术已经取得了较大的进展，并被证明可有效地加速材料研发-应用进程，因此被列为材料基因组计划的三大技



术要素之一[31]。“材料高通量实验”是在短时间内采用材料芯片等高通量方式完成大量样品的制备与表征。其核心思想是以并行处理替代传统材料研究中经常采用的顺序迭代尝试法，以高通量形式的量变引起材料研究方式的质变。高通量实验不仅可为高通量材料模拟计算提供海量的基础数据和实验验证；更为重要的是，高通量实验可快速地提供具有价值的研究成果，直接加速材料的筛选和优化。随着材料基因组计划的研究方法逐步被广泛采用，材料高通量制备技术的重要性将日益显现[31]。

随着高通量材料制备技术的推进，高通量材料表征和分析就应运而生，因为粉末衍射是材料晶体结构探测的经典方法，而晶体结构又是决定材料性能的至关重要因素，因此，高通量的粉末衍射自然成为高通量材料表征的重要一环。随着高通量粉末衍射产生的是粉末衍射数据规模的急剧扩展，传统的人工分析将面临极大地挑战，甚至已经无法应对如此大数据的汹涌而来。在这种情况下高通量的粉末衍射数据分析就变得非常急迫和重要。Rietveld 精修作为粉末衍射结构分析的重要方法，高通量的 Rietveld 精修自然而然就成为了研究的重要课题。因为 AutoFP 已经可以对单任务高度自动化地完成精修任务，这为高通量的精修创造了条件。

## 4.2 利用 AutoFP 实现高通量 Rietveld 精修

### 4.2.1 系统架构

系统架构如图 4-1 所示，系统整体采用 C/S 模型。高通量精修核心放在集群上，客户端可以通过网页提交任务，由 Sever 负责任务的解析和调度，继而由并行化的 AutoFP，MPI-AutoFP 负责求解任务，求解结果通过网页返回给用户。

**Client:** HTML5 网页，用户客户端，负责上传高通量精修任务，并展示精修结果。

**Website:** 采用 HTML5 Socket 技术，网站后台，1)负责与客户端通信，解析

客户端上传的高通量精修任务，并发送给高通量精修服务器 Server； 2)负责解析从 Server 传回来的精修结果，并发送给客户端显示精修结果。

Server: 1)高通量精修任务调度； 2)负责解析高通量精修任务给 MPI-AutoFP 执行，并在执行后收集信息，生成精修结果信息； 3)与 Analysis algorithm 交互数据，给算法分析提供条件，与 Other Software 交互数据，拓展系统功能。

MPI-AutoFP: 采用 MPI 并行化技术的 AutoFP 程序,可以在机群上并行化地执行大量精修任务。

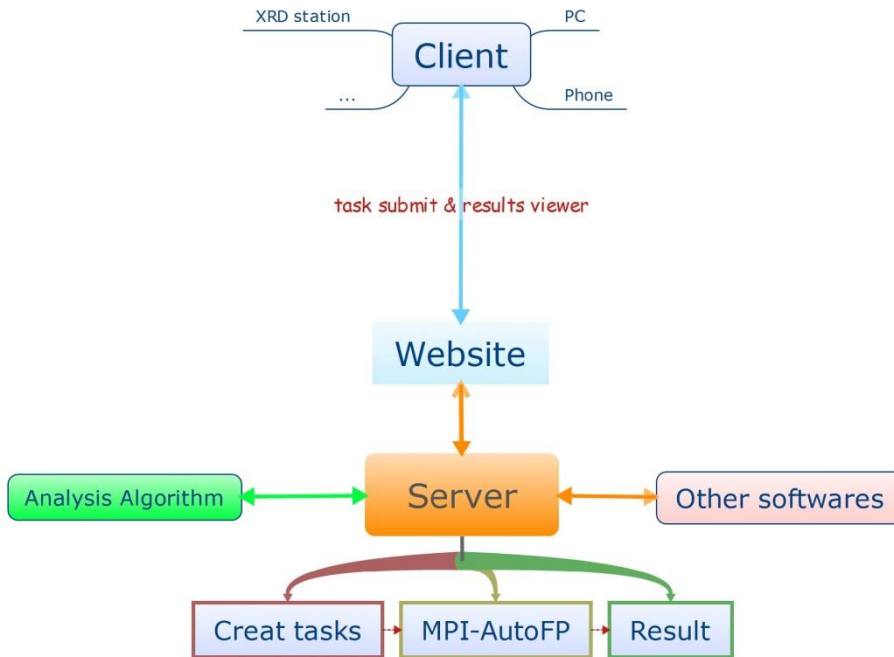


图 4 - 1 基于 AutoFP 的高通量 Rietveld 精修系统架构

Analysis Algorithm: 人工智能等算法,可以通过与 Sever 交互精修数据,对精修结果进行大数据分析,总结规律,做出预测。

Other softwares: 其他软件,比如 Material Studio, Vasp 等, 可以从 Server 获取高通量精修结果, 做其他方面的运算。

#### 4.2.2 AutoFP 程序 MPI 并行化

程序并行化的一种简单模型是多线程。多线程模型中存储空间是共享的，即线程之间可以互相访问，因此这一模型也仅仅能适用于共享的存储系统。然而实际情况中，并行机集群的各个处理器之间往往不能共享存储，这种情况下，运行于各个处理器的程序就必须通过来回传递消息的方式来通信。然而只有采用标准的消息传递库才能保证消息传递方式的可移植性，在这一要求下，并行化编程消息传递接口(即 Message Passing Interface)MPI 就应运而生。经过多年的发展，MPI 已经成为被广泛采用的并行编程消息传递标准，拥有 OpenMPI, MPICH2 等多种实现。

和基于共享存储模型的并行化机制 OpenMP 不同，MPI 是一种基于消息传递的并行化编程技术。值得注意的是 MPI 不是一种具体的编程语言,而只是一种并行编程消息传递接口标准。也就是说，MPI 标准事实上定义了一组具有可移植性的编程接口[45-46],在遵循 MPI 标准要求的条件下,用户可以实现自己需要的 MPI 软件包。因此 MPI 得多了多种实现，比如 MPICH、OpenMPI 等开源实现,以及商业公司 Intel 开发的 Intel MPI 不开源实现。在 MPI 统一并行编程接口的基础上,程序员可以专注于设计并行算法而不用担心具体的硬件平台,借助相应平台的 MPI 库即可实现并行计算和程序移植[47]。在 Linux, Unix, Windows 等多种主流操作系统上,MPI 都已经获得了支持,下面提到的上海大学自强 4000 平台主要并行化方案也是 MPI。

上海大学第三代集群式高性能计算机“自强 4000”于 2013 年 6 月建成，系统总运算峰值约 70 万亿次/秒（70Tflops），其中 CPU 双精度浮点运算峰值 57152Gflops，GPU 双精度浮点运算峰值 14410Gflops，系统总体 Linpack 效率约为 85%。

其主要包括：

- 5 个登录/管理节点; 4 个 I/O 节点; 1 套存储系统; 1 套 56Gb FDR Infiniband 计算网络; 1 套千兆以太网管理网络等。

- CPU 计算节点 140 台，每台 2 颗 Intel SandyBridge 架构的 E5-2690 CPU (2.9GHz/8-core)，64GB 内存，共有 2240 核、8.96TB 内存。
- 胖节点 2 台，每台 4 颗 Intel SandyBridge 架构的 E5-4650 CPU (2.7GHz/8-core)，512GB 内存，共有 64 核、1TB 内存。
- GPU 节点 11 台，每台 2 颗 Intel SandyBridge 架构的 E5-2680 CPU (2.7GHz/8-core)，64GB 内存，共计 176 核，每台 2 块 NVIDIA M2090 GPU 计算模块。

系统内存总量：10TB。

系统存储磁盘总量：180TB。

操作系统：Linux。

作业调度及系统管理与整合：Platform LSF。

并行文件系统：GPFS。

并行环境：Intel MPI、MPICH2 等 MPI 编译环境。

系统和集群监控管理：Intel Data Center Manager；IBM xCAT。

基于上海大学自强 4000 计算集群特征，MPI 是最适合的并行化方案。AutoFP 程序本身只能完成单个精修任务，为了适应高通量精修的需要，并行化的 AutoFP，MPI-AutoFP 被开发出来。MPI-AutoFP 采用 MPI 并行框架并发启动并调度多个 AutoFP 实例，并通过 JSON 数据交换格式收集所有 AutoFP 任务的工作状态和精修结果。AutoFP 本身具有优良的跨平台性能，所以 MPI-AutoFP 可以正常运行于基于 Linux 操作系统的上海大学超算集群上。

#### 4.2.3 高通量任务格式

高通量任务采用 zip.tar 等打包格式，包内文件组织，task.inf 为工程控制文件，用于指定任务类型和数据位置。

任务类型：

1. 多任务，每个 pcr 文件对应一个数据 dat 文件。这种模式适合于相互无关的大量任务集合并行化加速；

2. 单 pcr 文件对应多个 dat 文件。这种模式适合于单一结构，多数据的任务，比如对一个晶体连续降温所采集的 XRD 或者中子衍射，系统将一个 pcr 分发给全部数据，并行化启动全部任务，并对结果进行汇总展示；

3. 多个 pcr 文件对应单个 dat 文件。这种模式适合于一个数据文件，多个 pcr 结构模型的情况，比如对一个未知结构的晶体，可以采用大量尝试结构精修精修来寻找最可能的结构。

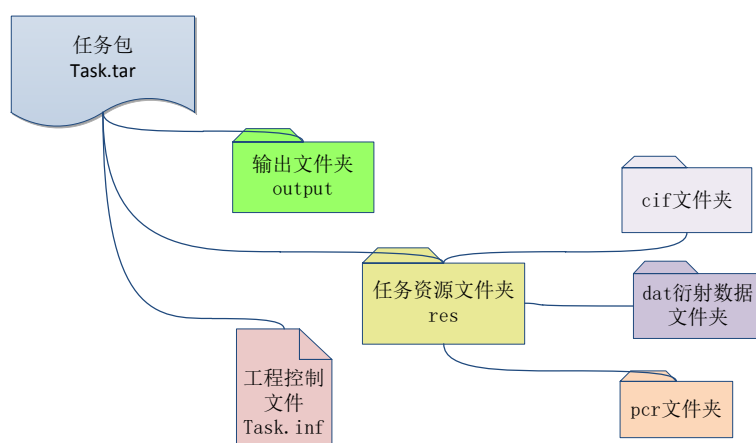


图 4 - 2 高通量 Rietveld 精修任务打包格式

#### 4. 2. 4 数据交换格式

高通量精修系统因为模块众多，而且涉及到不同的操作系统和硬件平台，因此各模块之间数据交换采用具有良好跨平台性的 JSON 数据交换格式。

JSON，全称 JavaScript Object Notation，是基于标准化脚本程序设计语言 ECMAScript 的一个子集的一种数据交换格式[48]。它具有使用广泛，轻量级，完全独立于特定编程语言等多种优点。JSON 数据以文本格式存储，数据结构使用了类似于 C/C++、Python 等 C 家族的习惯。这一系列特性使得 JSON 易于阅读，编写，解析，生成和传输(一般用于提升网络传输速率) [48]，成为高通量 Rietveld 精修系统数据交换的理想格式。

JSON 的规则非常简单：对象是一个无序的“‘键/值’对”集合。键/值对规则如下：

- 1、映射用冒号：表示。“键：值”
- 2、并列的数据之间用逗号，分隔。“键 1：值 1，键 2：值 2”
- 3、映射的集合（对象）用大括号（“{}”）表示。

```
{  
    键 1:值 1,  
    键 2:值 2  
}
```

- 4、并列数据的集合（数组）用方括号[]表示。

```
[  
    {键 1:值,键 2:值 2},  
    {键 1:值,键 2:值 2}  
]
```

- 5、元素值可具有的类型：string, number, object, array, true, false, null 等。

数据交换格式 JSON 主要有 2 种表示结构，即对象和数组。

对象结构以大括号开始结束。中间由多个以逗号分隔的”key(键): value(值)”对构成，语法结构如下。

```
{  
    key1:value1,  
    key2:value2,  
    ...  
}
```

其中键是字符串，而值可以是字符串，也可以是数值，或 true, false, null 等对象，甚至也可以为数组。

数组结构以中括号开始结束，中间由多个以逗号分隔的值列表组成，语法

结构如下所示。

```
[
  {
    key1:value1,
    key2:value2
  },
  {
    key3:value3,
    key4:value4
  }
]
```

在基于 AutoFP 的高通量精修系统中，各个任务的工作状态，精修参数和精修质量 $R_{wp}$ 等将以 JSON 格式存储并在各模块之间传递。表 4 - 1 展示了一个高通量 Rietveld 精修系统模块间交换数据的示例，该示例为 FullProf 精修后的 pcr 文件全部参数。各个模块都具备有 JSON 数据的编码和解码功能，为系统的模块化和模块的互操作性增加了便利。

表 4 - 1 高通量 Rietveld 精修系统数据交换格式实例

```
{ "Scale": 0.00038705, "Str3": 0.0, "Str2": 0.0, "Str1": 0.0,
  "BACK": 2.776, "Shape1": 0.0, "B12": 0.00036, "B13": 6e-05,
  "B11": -0.00061, "Sycos": 0.0, "B33": -7e-05, "D_L": 0.0,
  "Occ": 1.07465, "Pref1": 1.0, "alpha": 90.0, "PA2":
  -0.04941, "Bov": -4.25894, "Biso": 5.27994, "beta": 90.0,
  "PA3": 0.58223, "U": 0.010574, "W": 0.00577, "V": -0.019131,
  "Y": 0.045235, "X": 0.016963, "Pref2": 0.0, "Z": 0.38064,
  "Lambda": 0.0, "a": 10.603127, "PA1": -0.22087, "c":
  10.603127, "b": 10.603127, "PA4": 0.16472, "GausSiz":
  0.003432, "B23": 0.0008, "B22": 0.00057, "Zero": -0.04139,
  "LorSiz": 0.0, "Sysin": 0.04788, "S_L": 0.0, "gamma": 90.0 }
```

## 第五章 AutoFP 的综合应用测试实例

### 5.1 测试说明

晶体结构 Rietveld 精修往往需要很高的衍射数据质量，因此测试实例多采用常规 XRD 衍射仪以步进测量的方式获取衍射谱或者采用同步辐射等高亮度光源获取衍射谱，以提高衍射数据精度，达到测试需求。本章测试数据一部分由作者所在实验室合成并在上海大学测试中心测试，另一部分数据是 FullProf 自带的一些典型案例。

### 5.2 AutoFP 单相 Rietveld 精修测试

通过下面的测试例子可以展示 AutoFP 的功能和特性

#### 1) $Y_2O_3$

这个例子是文章 (C. Santos, 2005)[49]的主题，本例子展示了如何用 AutoFP 去完成一个简单的高对称性单相氧化物材料  $Y_2O_3$  的 Rietveld 精修。本例子所使用的数据是常规实验室 X 射线衍射仪所采集的。采集过程采用一个装备 Cu 靶的 Siemens d5000 衍射仪记录  $Y_2O_3$  多晶的高质量衍射谱。精修结果在图 5 - 3 和表 5 - 1 中展示，可以看出随着自动精修进程推进， $R_{wp}$  曲线不断下降到较小的值附近。而且，AutoFP 可以自动化地完成整个精修进程，可以解决精修过程中的发散情况和各种 FullProf 运行错误，使得最终的  $R_{wp}$  最小化。由图 5 - 1 和图 5 - 3 可见，AutoFP 的精修结果是非常完美的。本例证明了 AutoFP 可以高度自动化地完成简单无机材料粉末衍射的 Rietveld 精修任务，而且精修结果已经达到了很高的水平。



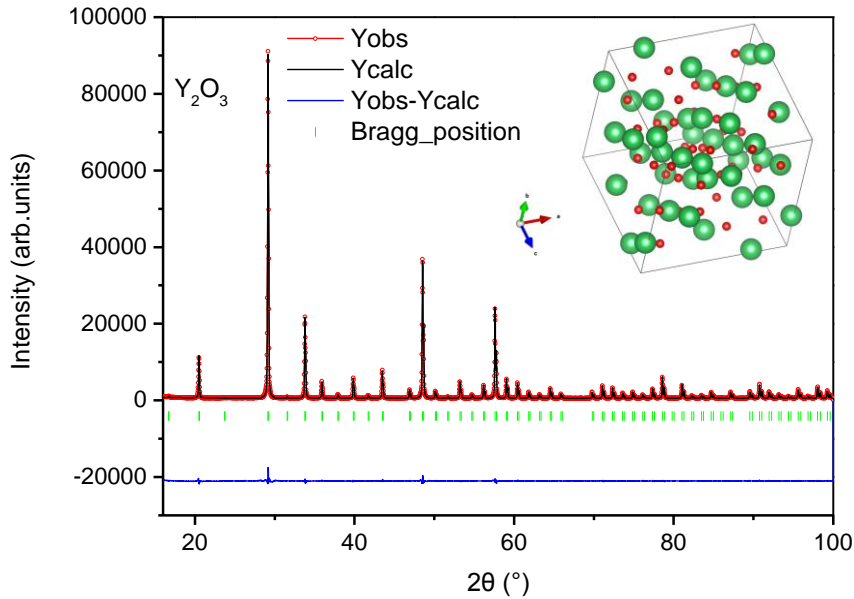


图 5 - 1 基于 AutoFP 的 Y2O3 精修结果图

表 5 - 1 基于 AutoFP 的 Y2O3 粉末数据 Rietveld 精修结果

Sample	Space group	Cell Unit(Å)	R-factors
Y <sub>2</sub> O <sub>3</sub>	Ia-3	a=b=c=10.60397(1)	R <sub>p</sub> = 6.74%, R <sub>wp</sub> = 7.23%, χ <sup>2</sup> = 2.69

## 2) Dy<sub>0.5</sub>Sr<sub>0.5</sub>MnO<sub>3</sub>

为了演示 AutoFP 对新物质自动化 Rietveld 精修的能力，我们采用 SrCO<sub>3</sub> (99.99%), Dy<sub>2</sub>O<sub>3</sub> (99.99%), MnO<sub>2</sub> (99.95%) 和标准固态反应法合成了 Dy<sub>0.5</sub>Sr<sub>0.5</sub>MnO<sub>3</sub> (DSMO) 粉末，DSMO 是具有多铁性的新材料，其晶体结构和多铁性关系具有重要研究价值[50]。在室温下使用一台“Rigaku 18 kW D/MAX 2550 X”射线粉末衍射仪 Cu 靶 Kα 线对该粉末测量了衍射谱。衍射仪在 40 kV 和 200mA 条件下操作，2θ 扫描范围为 10 到 120 度，步长为 0.02 度，4s 每步。

自动化 Rietveld 精修结果和 R<sub>wp</sub> 下降曲线展示在图 5 - 1 和图 5 - 3。因为 DSMO 是一种新合成的化合物。基于 La<sub>x</sub>Sr<sub>1-x</sub>MO<sub>3</sub> 的修改结构[6,51]被用来作为初始结构对 DSMO 进行精修。精修结果显示 DSMO 晶体结构空间群是 Pbnm，

晶格参数  $a=5.40196(4)$ ,  $b=5.41494(0)$  Å,  $c=7.62853(2)$  Å. 从这个例子可以看出 AutoFP 可以高效, 正确, 自动化地完成新结构的精修任务。这将极大促进新材料的研发和分析。

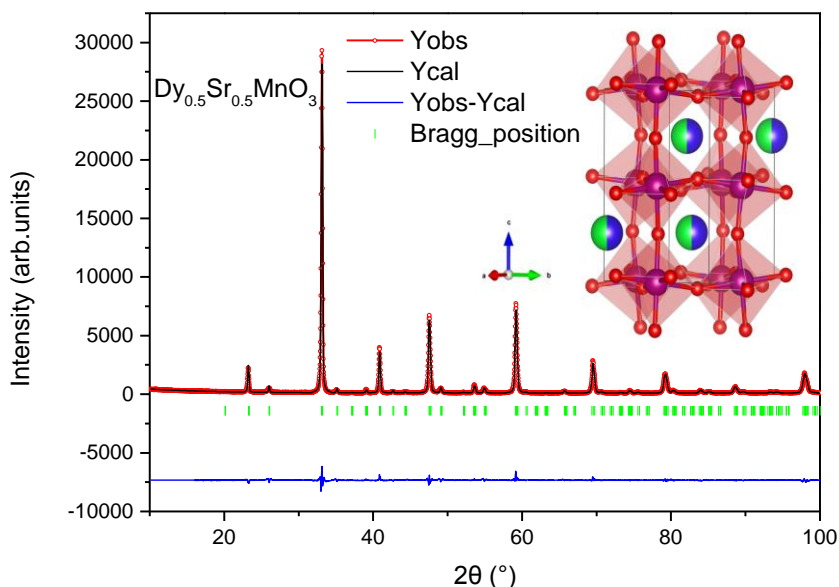


图 5-2 基于 AutoFP 的 DSMO 精修结果图

表 5-2 基于 AutoFP 的 DSMO 粉末衍射精修结果

Sample	Space group	Cell Unit(Å)	R-factors
$Dy_{0.5}Sr_{0.5}MnO_3$	$Pbnm$	$a=5.40196(4)$	$R_p=10.5\%$
		$b=5.41494(0)$	$R_{wp}=12.8\%$
		$c=7.62853(2)$	$\chi^2 = 3.56$

3)  $PbSO_4$ : 这个例子是文章 IUCr Round Robin[52]的主题。这是一个经典结构  $PbSO_4$  的定波长中子衍射数据。这个例子可以证明 AutoFP 可以自动化地完成典型的定波长中子衍射数据自动化精修任务。

4)  $LaMnO_3$ :  $LaMnO_3$  定波长中子粉末衍射[6]由 LLB 中子衍射仪 G4.2 用  $\lambda = 2.59$  Å 采集而成。

5)  $YBaCuCoO_7$ : 定波长中子粉末衍射数据。该数据从 FPschool2013 网站

获取，由 I.L.L. 使用仪器 D1A 对一个氧化物样品  $YBa_2Cu_{3-x}CO_xO_{7+y}$  在  $x = 0.33$  条件下测量所得。

表 5-3 AutoFP 精修实例精修结果

Sample	$R_p(\%)$	$R_{wp}(\%)$	$\chi^2$	Diffraction Mode
$Y_2O_3$	7.27	6.74	2.69	Xray
$PbSO_4$	4.99	5.78	3.63	Neutron-CW
$DSMO$	12.7	10.5	3.56	Xray
$LaMnO_3$	11.7	15.0	1.40	Neutron-CW
$YBaCuCoO_7$	7.64	7.31	1.59	Neutron-CW

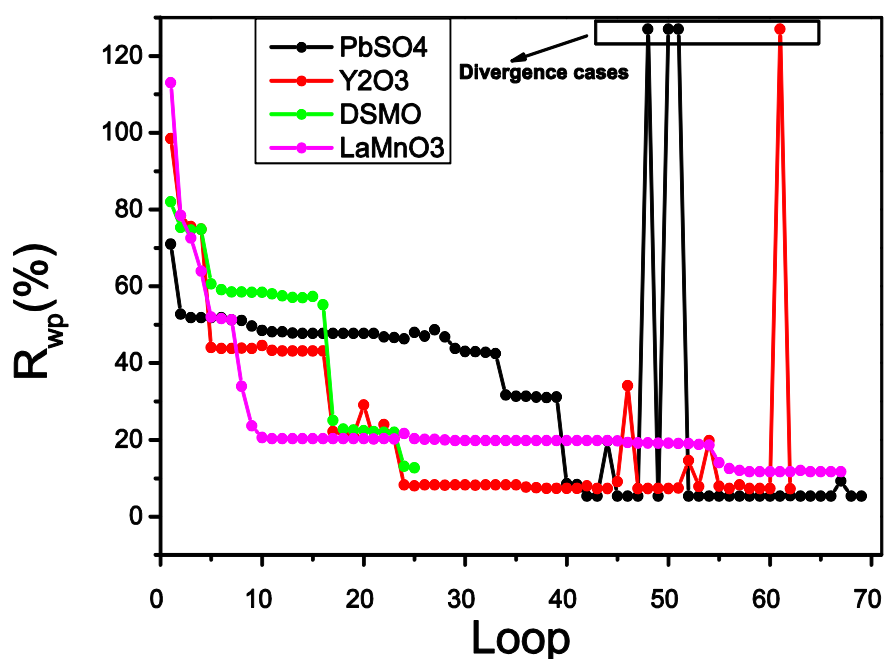


图 5-3  $R_{wp}$  下降曲线，由 AutoFP 自动化精修  $PbSO_4$ ,  $Y_2O_3$ ,  $DSMO$  和  $LaMnO_3$  等示例所生成

### 5.3 基于 AutoFP 的自动化定量相分析测试

例子 1) 本例子是混合制备出来的  $CaVO_3$  和  $CaV_2O_4$  混合物，可以采用 AutoFP 对这样的混合物进行定量相分析，从图 5-4 可以看出自动化定量相分析

结果比较理想，误差很低，结果显示  $\text{CaVO}_3$  相和  $\text{CaV}_2\text{O}_4$  相的质量分数分别是 61%，39%。该例子证明 AutoFP 可以很好地完成两相的定量相分析精修任务，同时也展示了 AutoFP 自带工具 prf2origin 的自动化画图能力。

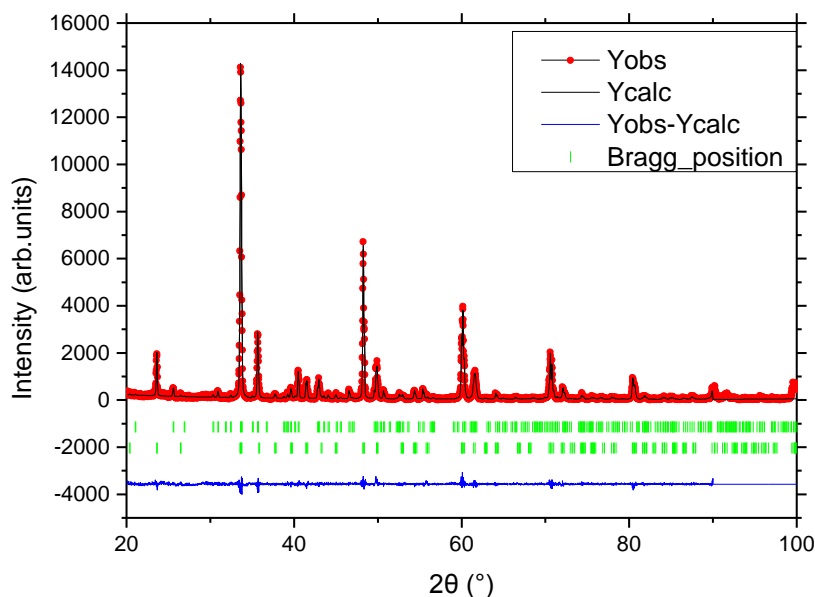


图 5-4 基于 AutoFP 的  $\text{CaVO}_3$  和  $\text{CaV}_2\text{O}_4$  混合物自动化定量多相 Rietveld 精修结果。其中 Bragg\_position 上排为  $\text{CaV}_2\text{O}_4$  峰，下排为  $\text{CaVO}_3$  峰。本图由 AutoFP 所集成的 prf2origin 工具自动生成。

## 5.4 基于 AutoFP 的高通量 Rietveld 精修系统测试

例子 1) 高通量 Rietveld 精修流程演示，随机选取 32 个单相精修任务，然后组成一个高通量 Rietveld 精修任务包，提交任务包图 5-5，启动高通量 Rietveld 精修，可以看出基于 AutoFP 的高通量 Rietveld 精修系统采用并行模式，快速地完成了这 32 个精修任务，并返回  $R_{wp}$  曲线图 5-8，和晶体结构晶胞参数列表图 5-7。证明了基于 AutoFP 的高通量 Rietveld 精修系统有能力完成大规模衍射数据的高通量分析。

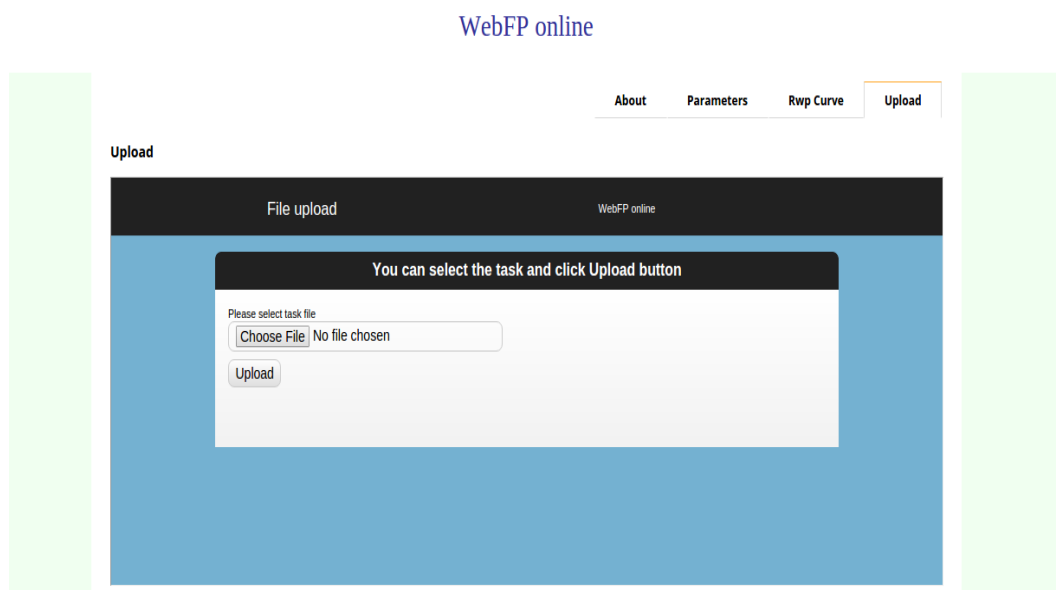


图 5 - 5 高通量精修系统 上传任务



图 5 - 6 高通量精修系统 任务包

### WebFP online

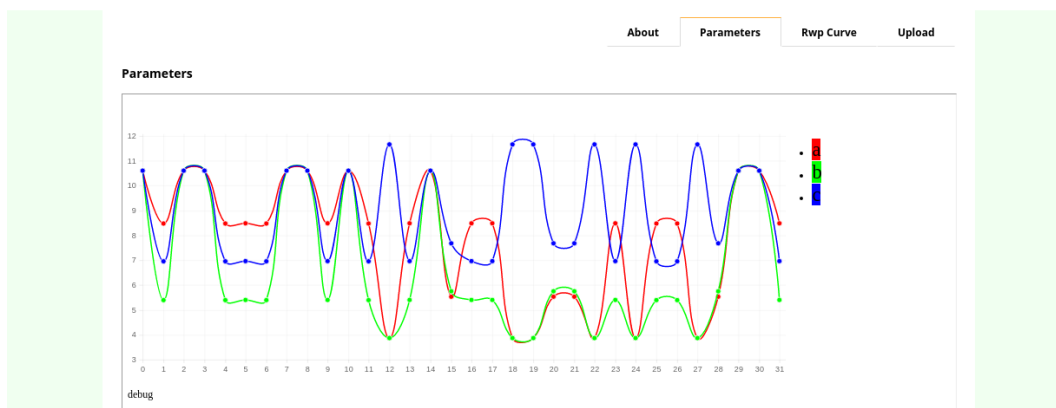


图 5-7 高通量精修系统 任务结果 (晶体结构)

### WebFP online

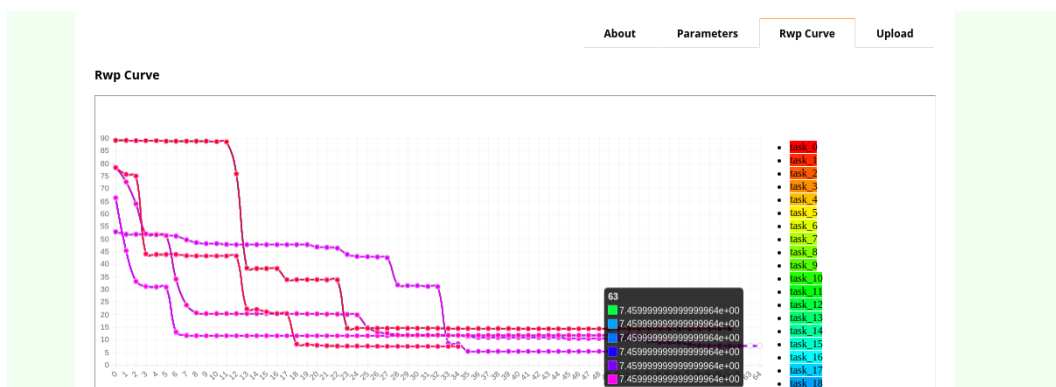


图 5-8 高通量精修系统 任务结果 ( $R_{wp}$ )

## 第六章 结论与展望

### 6.1 结论

1. Rietveld 精修是粉末衍射分析的重要方法，但传统 Rietveld 精修程序普遍存在着效率低，操作复杂的缺点。本文基于人工智能算法和广泛使用的 Rietveld 精修程序 FullProf 开发了一款高度自动化的 Rietveld 精修工具集 AutoFP。AutoFP 的开发极大提高了 Rietveld 精修程序 FullProf 的自动化水平。它采用专家系统的算法控制 FullProf 去模拟人类专家的精修过程，从而让程序可以高度自动化和智能化地完成复杂的精修任务。由此证明了专家系统是实现 Rietveld 精修自动化的很好选择。同时，AutoFP 提供了基于机器学习算法的 Rietveld 精修策略搜索模块来提升专家系统策略库，AutoFP 也为高级用户保留了编程接口，方便了他们设计自己的精修策略，或者使用新的精修引擎。AutoFP 也可以被其他程序作为一个自动化的精修引擎所调用。AutoFP 是采用 Python 代码的开源工程，用户界面友好，容易学习和使用。

2. 基于 AutoFP 的 Rietveld 精修自动化能力，本文继续开发了基于 AutoFP 的自动化定量相分析程序，大大简化了采用 FullProf 进行定量相分析的操作。

3. 随着材料基因组计划的提出，高通量的材料表征成为重点，其中，高通量粉末衍射将产生大量的粉末衍射数据，面对如此多的数据，高通量的粉末衍射数据分析就变得非常迫切，本文基于自动化 Rietveld 精修程序 AutoFP 开发了高通量的 Rietveld 精修系统，为高通量 Rietveld 精修创造了可能性。

4. 测试表明，自动化 Rietveld 精修程序 AutoFP 可以高度自动化地完成单相 Rietveld 精修任务，且结果非常理想。对 AutoFP 改进而实现的自动化定量相分析，极大地简化了 FullProf 定量相分析的操作，可以迅速完成定量相分析精修任务，并自动化给出定量相分析结果。基于 AutoFP 的高通量 Rietveld 精修系统有能力实现大规模衍射数据的高通量分析，有利于推动高通量材料表征和分析技术的发展，进而推动材料基因组工程的实施。

## 6.2 展望

通过对编写基于人工智能算法的自动化和高通量的 Rietveld 精修程序 AutoFP 的探索以及对 AutoFP 自动化，高通量 Rietveld 精修的测试，作者认为该研究领域还存在以下几方面值得进一步扩充和深入：

1. 基于机器学习算法的 Rietveld 精修策略搜索模块仍有大量完善和拓展的空间；更多其他的机器学习算法可以加以尝试。机器学习算法的性能发挥依赖于海量的练习数据[3,53]，建设供 Rietveld 精修机器学习的粉末衍射数据库也势在必行。

2. 中子衍射中，磁结构的 Rietveld 精修自动化解析在 AutoFP 中尚未完全实现自动化。中子具有自旋，能与电子自旋发生相互作用，中子衍射可以反应固体中电子自旋的排列，而电子自旋排列对磁学和自旋电子学具有重要意义。因此，基于中子衍射的磁结构和电子自旋排列分析对凝聚态研究具有重要价值，完整实现含有磁结构的中子衍射 Rietveld 精修自动化有重要现实意义。

3. 高通量 Rietveld 精修平台需要进一步完善。高通量材料制备技术已经逐渐起步，而高通量材料表征，特别高通量粉末衍射仍然有很多困难[31]，而高通量 Rietveld 精修系统的建设和改进有赖于高通量衍射数据的获取。另外，高通量精修结果的大数据分析对材料结构的规律总结和预测研究具有重要意义，所以高通量精修结果的大数据分析技术的研究也势在必行。此外，高通量 Rietveld 精修和其他材料计算软件（比如 MS，VASP 等）的数据交互对高通量材料计算模拟有着重要价值，这方面也有待提高。



## 参考文献

- [1] X.P. Cui, Z.J. Feng, Y. Jin, et al., “AutoFP: a GUI for highly automated Rietveld refinement using an expert system algorithm based on FullProf” [J]. *J. Appl. Cryst.*, Vol. 48, No.5, 2015, pp. 1581~1586
- [2] P. Tian, W. Zhou, J. Liu, et al., “SrRietveld: a program for automating Rietveld refinements for high-throughput powder diffraction studies” [J]. *J. Appl. Cryst.*, Vol. 46, No. 1, 2013, pp. 255~258
- [3] D. Silver, A. Huang, et al., “Mastering the game of Go with deep neural networks and tree search”[J]. *Nature*, Vol.529, No.7587, 2016, p.484
- [4] A. Le Bail, “A whole powder pattern decomposition methods and applications: a retrospection” [J]. *Powder Diffraction* , Vol.20, No.4, 2005, pp.316~326
- [5] E. F. Bertaut, “Spin Configurations in Ionic Structures: Theory and Practice, in Magnetism” [J]. Vol.3, Ed. G.T. Rado and H. Suhl, Academic Press, 1963
- [6] J. Rodriguez-Carvajal, J. guez-Carvajal, M. Hennion, et al., “Neutron-diffraction study of the Jahn-Teller transition in stoichiometric LaMnO<sub>3</sub>” [J]. *Phys. Rev. B*, Vol.57, No.6, 1998, pp.3189~3192
- [7] M. Burrard-Lucas, D. G. Free, et al., “Enhancement of the superconducting transition temperature of FeSe by intercalation of a molecular spacer layer”[J]. *Nature Materials*, Vol.12, No.1, 2013, pp.15~19
- [8] 梁敬魁, “粉末衍射法测定晶体结构(上,下)” [M]. 科学出版社, 2003.
- [9] W. I. F. David, K. Shankland, “Structure determination from powder diffraction data” [J]. *Acta Cryst. A*, Vol. 64, No. 1, 2008, pp. 52~64
- [10] 马礼敦, “近代 X 射线多晶体衍射: 实验技术与数据分析” [M]. 北京,化学

工业出版社, 2004.

- [11] 张海军, 贾全利, 董林编, “粉末多晶 X 射线衍射技术原理及应用” [M]. 郑州大学出版社, 2010, pp.17~226
- [12] 冯振杰, “遗传算法用于粉末衍射全谱拟合与晶体结构测定” [J]. 博士学位论文, 2007.
- [13] H. M. Rietveld, “A profile refinement method for nuclear and magnetic structures” [J]. *J. Appl. Cryst.*, Vol. 2, No. 2, 1969, pp. 65~71
- [14] H. M. Rietveld, “Line profiles of neutron powder-diffraction peaks for structure refinement” [J]. *Acta Cryst.*, Vol. 22, No. 1, 1967, pp. 151~152
- [15] Z. J. Feng, C. Dong, “GEST: a program for structure determination from powder diffraction data using a genetic algorithm” [J]. *J. Appl. Cryst.*, Vol. 40, No. 3, 2007, pp. 583~588
- [16] Z. J. Feng, C. Dong, R. R. Jia, X. D. Deng, S. X. Cao, J. C. Zhang, “PeckCryst: a program for structure determination from powder diffraction data using a particle swarm optimization algorithm” [J]. *J. Appl. Cryst.*, Vol. 42, No. 6, 2009, pp. 1189~1193
- [17] Z. J. Feng, Y. L. Zheng, Y. L. Su, B. Shao, M. T. Li, S. X. Cao, J. C. Zhang, “New procedure to obtain Bragg-reflection intensities from FULLPROF suite for powder crystal-structure determination using GEST and PECKCRYST programs” [J]. *Powder Diffraction*, Vol. 25, No. 04, 2012, pp. 374~376
- [18] G. Calestani, C. Rizzoli, “Crystal structure of the  $\text{YBa}_2\text{Cu}_3\text{O}_7$  superconductor by single-crystal X-ray diffraction” [J]. *Nature*, Vol.328, No.6131, 1987, pp.606~607
- [19] R. A. Young, A. Sakthivel, T. S. Moss, C. O. Paiva-Santos, “DBWS-9411-an upgrade of the DBWS\*.\* programs for Rietveld refinement with PC and

- mainframe computers” [J]. *J. Appl. Cryst.*, Vol. 28, No. 3, 1995, pp. 366~367
- [20] J. Rodriguez-Carvajal, “FULLPROF: A Program for Rietveld Refinement and Pattern Matching Analysis” [J]. Abstracts of the Satellite Meeting on Powder Diffraction of the XV Congress of the IUCr, Toulouse, France, 1990, p. 127
- [21] A. C. Larson, R. B. Von Dreele, “General Structure Analysis System (GSAS)” [J]. Los Alamos National Lab Rep. LAUR, 1994, pp. 86~748
- [22] B. H. Toby, “EXPGUI, a graphical user interface for GSAS” [J]. *J. Appl. Cryst.*, Vol. 34, No. 2, 2001, pp. 210~213
- [23] F. Izumi, T. Ikeda, “Rietan2000” [J]. *Mater. Sci. Forum*, Vol. 321-324, 2000, pp. 198~203
- [24] C. Dong, “PowderX: Windows-95-based program for powder X-ray diffraction data processing” [J]. *J. Appl. Cryst.*, Vol. 32, 1999, p. 838
- [25] SH, Liao, “Expert system methodologies and applications - a decade review from 1995 to 2004” [J]. *Expert Syst. Appl.*, Vol. 28, No.1, 2005, pp. 93~103
- [26] M. Wooldridge, N.R. Jennings, “Intelligent agents: theory and practice” [J]. *Knowl. Eng. Rev.*, Vol.10, No.2, 1995, pp.115~152
- [27] M. Pantic, LJM Rothkrantz, “Expert system for automatic analysis of facial expressions” [J]. *Image Vis. Comput.*, Vol.18, No.11, 2000, pp.881~905
- [28] C. Watkins, P. Dayan, “Learning from delayed rewards. 1989” [J]. Cambridge University, Cambridge, 1966
- [29] H. W. CJC, “Learning from delayed rewards” [J]. Cambridge University, Cambridge, England, Doctoral thesis, 1989
- [30] 刘晓轩, “Rietveld 方法在无机材料中的一些应用” [J]. 硕士学位论文, 2006, pp.12~15

- [31] 王海舟, 汪洪, 丁洪, 项晓东, 向勇, 张晓琨, “材料的高通量制备与表征技术” [J]. 科技导报, Vol. 33, No. 10, 2015, pp.31~46
- [32] 赵继成, “材料基因组计划简介”[J]. Chinese Journal of Nature, Vol.36, No.2, 2014, p.89
- [33] P. Tian, S. Billinge, “Testing different methods for estimating uncertainties on Rietveld refined parameters using SrRietveld” [J]. Zeitschrift für Kristallographie, Vol. 226, No. 12, 2011, pp. 898~904
- [34] C. Giacovazzo, H. L. Monaco, D. Viterbo, F. Scordari, G. Gilli, G. Zanotti, M. Catti, “Fundamentals of Crystallography. International Union of Crystallography” [J]. Oxford University Press, USA, 2 edition, 2002.
- [35] MF Sanner, “Python: A programming language for software integration and development” [J]. J. Mol. Graph. Vol.17, No.1, 1999, pp.57~61
- [36] B. Rempt, “Python’s PyQt toolkit” [J]. Dr. Dobbs J., Vol.26, No.1, 2001, p.88
- [37] H. Marcus, “Universal Artificial Intelligence”[M]. Springer, 2005.
- [38] C. T. Leondes, “Expert Systems: The Technology of Knowledge Management and Decision Making for the 21st Century”[M]. Academic Press, 2002.
- [39] R. A. Young, “The Rietveld Method”[M]. Oxford University Press, Vol.5, 1993.
- [40] L. B. McCusker, R. B. Von Dreele, et al., “Rietveld refinement guidelines”. J. Appl. Cryst., Vol.32, 1999, pp.36~50
- [41] P. Bille, M. Farach-Colton, “Fast and compact regular expression matching”[J]. Theoretical Computer Science, Vol.409, No.3, 2008, pp.486~496
- [42] A. Moreo, E. M. Eisman, et al., “Learning regular expressions to template-based FAQ retrieval systems”[J]. Knowledge-Based Systems, Vol.53, 2013, pp.108~128

- [43] 褚建华, “Q-learning 强化学习算法改进及其应用研究” [J]. 北京化工大学, 2009.
- [44] Watkins, C. J. C. H., Dayan, P., “Q-learning”[J]. Machine Learning, Vol.8, No.3, 1992, pp.92~279
- [45] N. T. Karonis, B. Toonen, I. Foster, “MPICH-G2: A Grid-enabled implementation of the Message Passing Interface”[J]. Journal of Parallel and Distributed Computing, Vol.63, No.5, 2003, pp.551~563
- [46] N. F. Chilton, R. P. Anderson, et al., “PHI: A powerful new program for the analysis of anisotropic monomeric and exchange-coupled polynuclear d- and f-block complexes”[J]. Journal of Computational Chemistry, Vol.34, No.13, 2013, pp.1164~1175
- [47] D. C. Lonie, E. Zurek “XTALOPT: An open-source evolutionary algorithm for crystal structure prediction”[J]. Computer Physics Communications, Vol.182, No.2, 2011, pp.372~387
- [48] J. J. Merelo-Guervos, P. A. Castillo, et al., “Asynchronous Distributed Genetic Algorithms with Javascript and JSON”[J]. 2008 Ieee Congress on Evolutionary Computation, Vol.1-8, 2008, pp.1372~1379
- [49] C. Santos, K. Strecker, et al., “Stabilization of  $\alpha$ -SiAlONs using a rare-earth mixed oxide (RE<sub>2</sub>O<sub>3</sub>) as sintering additive” [J]. Materials Research Bulletin, Vol.40, No.7, 2005, pp. 1094 ~1103
- [50] Y. Jin, X.P. Cui, et al., “Spin reversal and ferroelectricity in perovskite Dy<sub>0.7</sub>Sr<sub>0.3</sub>MnO<sub>3</sub> and Dy<sub>0.6</sub>Sr<sub>0.4</sub>MnO<sub>3</sub>” [J]. Applied Physics Letters, Vol. 107, No.7, 2015, p.072907
- [51] L. Pinsard, J. Rodriguez-Carvajal, A. Revcolevschi, “Structural phase diagram of La<sub>1-x</sub>Sr<sub>x</sub>MnO<sub>3</sub> for low Sr doping”[J]. Journal of Alloys and Compounds, Vol.262, 1997, pp.152~156

- [52] R. J. Hill, "Rietveld refinement round robin. I. Analysis of standard X-ray and neutron data for PbSO<sub>4</sub>" [J]. *J. Appl. Cryst.*, Vol.25, No.5, 1992, pp.589~610
- [53] L. Shane, H. Marcus, "Universal Intelligence: A Definition of Machine Intelligence"[J]. Vol.17, No.4, 2007, pp.391~444

## 作者在攻读硕士学位期间公开发表的论文

- [1] **Xiaopeng Cui**, Zhenjie Feng, Yuan Jin, Yiming Cao, Dongmei Deng, Hao Chu, Shixun Cao, Cheng Dong, Jincang Zhang, “AutoFP: a GUI for highly automated Rietveld refinement using an expert system algorithm based on FullProf” [J]. *Journal of Applied Crystallography*, Vol. 48, No.5, 2015, pp. 1581~1586
- [2] Yuan Jin, **Xiaopeng Cui**, Weihua Han, Shixun Cao, Yuze Gao, Jincang Zhang, “Influence of the interface in quantum corrections to low-temperature resistance of  $\text{La}_{2/3}\text{Sr}_{1/3}\text{MnO}_3$  trilayer masking thin films” [J]. *Physical Chemistry Chemical Physics*, Vol.17, No.19, 2015, pp. 12826~12832
- [3] Yuan Jin, **Xiaopeng Cui**, Shixun Cao, Juanao Cheng, Wei Ren, Jincang Zhang, “Spin reversal and ferroelectricity in perovskite  $\text{Dy}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$  and  $\text{Dy}_{0.6}\text{Sr}_{0.4}\text{MnO}_3$ ” [J]. *Applied Physics Letters*, Vol. 107, No.7, 2015, p.072907
- [4] Yuan Jin, **Xiaopeng Cui**, Yuze Gao, Yifei Fang, Bo Lu, Shixun Cao, Jincang Zhang, “Temperature-induced variation of magnetoresistance and tunneling current in  $\text{La}_{2/3}\text{Sr}_{1/3}\text{MnO}_3/\text{LaMnO}_3(\text{ZrO}_2)/\text{La}_{2/3}\text{Sr}_{1/3}\text{MnO}_3$  junctions” [J]. *Solid State Communications*, 2015, Vol.215, pp. 12~16

## 作者在攻读硕士学位期间所承担的项目

- [1] 国家自然科学基金: “强关联氧化物体系的应力调制及多场耦合效应”, 直接费用 73.00 万元(No.11574194)
- [2] 上海市浦江人才计划项目: “新材料晶体结构 Rietveld 精修自动化探索”(No.13PJD015)
- [3] 上海市教育委员会科研创新重点项目: “强关联氧化物超导体的应力调制效应与量子相干输运机制研究”, 2012 年(No.12ZZ097)



## 致 谢

本文是在导师张金仓教授的悉心指导下完成的。承蒙张金仓老师的亲切关怀和精心指导，虽然有繁重的科研和管理工作，但仍抽出时间给予我学术上的指导和帮助，特别是给我提供了良好的科研和学习条件以及生活上的无微不至的关心，使我获益菲浅，并深受感激。其中，自动化和高通量 Rietveld 精修深受张老师材料基因组工程的启发和推动。张老师对学生认真负责的态度、严谨的科学研究方法、敏锐的学术洞察力、勤勉的工作作风以及勇于创新、勇于开拓的精神是我们永远学习的榜样。在此，谨向张老师致以深深的敬意和由衷的感谢。

感谢曹世勋教授对程序功能给予的意见和建议。感谢冯振杰副教授在 Rietveld 精修学习方面提供的入门引导和其新材料晶体结构 Rietveld 精修自动化项目的早期探索和支持。感谢郑咏雷同学在 Q 学习算法用于 Rietveld 精修的早期探索。

感谢金源师兄在多个实验和文章上的合作，以及给予的鼓励，帮助和支持，感谢钱小龙，方依菲等学长和卢霄文同学，以及黎杰等学弟们给与的鼓励和帮助以及实验上的大力支持，感谢郭沛音同学在定量相分析测试中的实验支持。

感谢上海大学自强 4000 超算集群提供的计算支持。

还要感谢我的家人，他们在生活上给予我很大的支持和鼓励，是他们给予我努力学习的信心和力量。

最后，感谢所有关心我、支持我和帮助过我的同学、朋友、老师和亲人。在这里，我仅用一句话来表明我无法言语的心情：感谢你们！