

中国科学技术大学

专业硕士学位论文

(工程硕士)



基于深度学习的 RAFM 钢晶界氦效 应研究

作者姓名: 陈 村
学科专业: 核科学与核技术
导师姓名: 彭 蕾 副教授
企业导师: 吴学邦 研究员
完成时间: 二〇二一年五月

University of Science and Technology of China

A dissertation for master's degree

(Master of Engineering)



Research on Helium Effects of Grain Boundary in RAFM Steel Based on Deep Learning

Author's Name: Cun Chen

Speciality: Nuclear Energy Science and Engineering

Supervisor: Associate Prof. Lei Peng

Research: Prof. Xuebang Wu

Finished time: May, 2021

摘 要

近年来,低活化铁素体马氏体(RAFM)钢作为一种候选的核结构材料得到了飞速的发展。RAFM 钢的性能在受到高能中子辐照后会因为辐照氦效应而有所下降。辐照产生的氦与 RAFM 钢中的位错、晶界等缺陷相互发生作用,形成纳米量级的氦泡,从而影响材料的力学性能。因此,研究 RAFM 钢中的晶界氦效应是十分有意义的。本文通过深度学习方法,针对该过程中的晶界结构与宏观性能间的关联性问题以及挖掘间隙氦的微观因素对迁移的影响,分别建立了对应的模型并进行了结果分析。

快速获得各种类型晶界性质是研究晶界氦效应的前提,本文基于 TextCNN 模型预测了 α -Fe 对称倾斜晶界的晶界能。首先通过分子动力学构建了 α -Fe 中倾斜轴为 $\langle 110 \rangle$ 和 $\langle 100 \rangle$ 的对称倾斜晶界库,再以晶界分布提取出径向分布函数、中心对称参数和积累中心对称参数三种晶界结构特征,最后分别通过这三种特征训练了 TextCNN 预测晶界能模型。在测试集上评价了各特征参数预测晶界能的平均相对误差,其中径向分布函数特征为 5.15%,中心对称参数 4.17%,积累中心对称参数为 1.74%,平均每个晶界的预测用时在 0.002s 以内。可见,积累中心对称参数作为晶界结构特征能在 TextCNN 模型下建立与晶界能间最紧密的连接。研究过程中,通过按倾斜轴分层抽样和数据增强的方法提升了模型的效果。

同时,本文还利用 LSTM 网络构建了判断间隙氦原子在晶界中是否迁移的分类模型,并在 α -Fe 的完整晶粒和 $\langle 110 \rangle$ 倾斜轴的 $\Sigma 3\{112\}$ 和 $\Sigma 11\{113\}$ 对称倾斜晶界模型中进行了应用。在测试集上进行评价,完整晶粒模型氦迁移的查准率为 0.82、查全率为 0.83、F1 分数为 0.82; $\Sigma 3$ 晶界模型氦迁移的查准率为 0.82、查全率为 0.87、F1 分数为 0.84; $\Sigma 11$ 晶界模型氦迁移的查准率为 0.82、查全率为 0.85、F1 分数为 0.83。同时,分析了氦原子的总速度大小、速度分量、速度方向夹角和铁氦结构参数特征与间隙氦在晶界中的迁移之间的关联性。初步的结果有:间隙氦所处环境(铁氦结构参数)对判断是否迁移具有重要的作用,同时,间隙氦的速度特征较环境特征对判断迁移更为重要。速度特征的冗余可能对模型造成负面影响,但不能完全没有速度特征。同一温度(350K)下,模型在 $\Sigma 11\{113\}$ 晶界(1113mJ/m²)上对速度方向夹角和铁氦结构参数的敏感度远高于完整晶粒和 $\Sigma 3\{112\}$ 晶界(260mJ/m²)。

关键词: α -Fe; 晶界; 深度学习; 晶界能; 间隙氦; 中心对称参数

ABSTRACT

In recent years, low activation ferritic martensitic steel as a candidate nuclear structure material has been rapidly developed. However, the performance of RAFM steel will be reduced after high-energy neutrons irradiating because of the irradiation helium effect. Meanwhile, the dislocations, grain boundaries and other defects in RAFM will interact with the helium. Therefore, it's significant to study the relationship between the grain boundary microstructure and macroscopic properties in RAFM steel. In this paper, the deep learning method was used to study the correlation between the grain boundary energy and the α -Fe symmetric tilt grain boundaries generated by molecular dynamics, and we also study the correlation between the migration probability of the He interstitials and the microscopic information of helium atoms

In this paper, TextCNN model is used to predict the grain boundary energy of the α -Fe symmetric tilt grain boundaries. Firstly, we constructed the α -Fe symmetric tilt grain boundary database with the tilt axis of $\langle 100 \rangle$ and $\langle 110 \rangle$. Then, three grain boundary structure features (RDF, CS, ACS) were extracted from the atomic distribution. Finally, the TextCNN model for predicting grain boundary energy was trained based on these three features. The average relative error of predicting grain boundary energy was evaluated on the test set, which the RDF was 5.15%, the CS was 4.17% and the ACS was 1.74%. The average prediction time of each grain boundary was less than 0.002s. It can be seen that the accumulated centro-symmetry parameters as the TextCNN feature can establish the closest connection with the grain boundaries. The methods of data augmentation and stratified sampling according to the tilt angle was proposed to improve the capability of the model.

In addition, the classification model of the He interstitials migration in α -Fe grain boundary ($\langle 110 \rangle \Sigma 3 \{112\}$ and $\langle 110 \rangle \Sigma 11 \{113\}$) was constructed by LSTM network. The precision, recall and F1 score of the complete grain model was 0.82, 0.83 and 0.82 respectively. The precision, recall and F1 score of the $\Sigma 3$ grain boundary model was 0.82, 0.87 and 0.84 respectively. The precision, recall and F1 score of the $\Sigma 11$ grain boundary model was 0.82, 0.85 and 0.83 respectively. We analyzed the correlation between the He interstitials migration and the helium velocity, velocity component, velocity direction, centro-symmetric parameters. The CS has more obvious influence on He interstitials migration, and with at the high temperature, the influence of the

helium velocity decreases.

Key Words: α -Fe; Grain boundary; Deep learning; Grain boundary energy; He interstitials; Centro-symmetry parameters

目 录

第 1 章 绪论.....	1
1.1 低活化铁素体/马氏体钢.....	1
1.2 辐照氦效应多尺度模拟方法.....	2
1.3 对称倾斜晶界/晶界能.....	3
1.4 机器学习/深度学习概述.....	5
1.5 本文研究意义和主要内容.....	6
第 2 章 机器学习/深度学习方法原理.....	8
2.1 支持向量机.....	8
2.2 卷积神经网络.....	9
2.3 循环神经网络.....	14
2.4 本章小结.....	15
第 3 章 基于卷积神经网络研究晶界能和晶界结构相关性.....	16
3.1 建立 α -Fe 对称倾斜晶界库.....	16
3.2 提取晶界结构特征.....	20
3.3 设计 TextCNN 模型.....	26
3.4 预测晶界能结果.....	30
3.5 本章小结.....	38
第 4 章 基于长短时记忆网络研究间隙氦在晶界中的迁移.....	39
4.1 建立间隙氦迁移样本库.....	39
4.2 提取氦原子迁移特征参数.....	41
4.3 设计 LSTM 模型.....	42
4.4 模型结果分析.....	44
4.5 本章小结.....	50
第 5 章 总结与展望.....	51
5.1 总结.....	51
5.2 展望.....	51
参考文献.....	52

第 1 章 绪论

1.1 低活化铁素体/马氏体钢

在聚变堆中，第一壁等结构材料需要承受高通量的高能中子辐照^[1]。因此，材料的选择是十分严苛的，相比于工业成熟发展的奥氏体不锈钢，低活化铁素体/马氏体钢(RAFM 钢)^[2]在强辐照下具有热导率高、抗辐照肿胀能力强和低活化的优点。同样的候选材料还有钒合金、SiCf/SiC 复合材料等，相比于它们，RAFM 钢在制造工艺和性能研究上均有更深的积累。其发展历程如下：1983 年，Bloom E F 等^[3]在体心结构的基础上，通过加入低活性的元素得到了能减弱中子辐照效应的低活化材料；国际能源组织在低活化材料的概念上于 1992 年进行了低活化马氏体钢的研究工作^[4]。目前，有美国橡树岭国家实验室(ORNL)开发的 9Cr-2WVTa，日本原子能源机构和 JFE 钢铁公司开发的 F82H，欧洲 FZK 和 CEA 联合开发的 EUROFER97 以及中国科学院等离子体物理研究所开发的 CLAM 等^[5]，主要成分如下表 1.1 所示。想要直接模拟完整的 RAFM 钢结构具有一定的难度，因此本文通过体心立方结构(Body centered cubic, BCC)的 α -Fe 做简化的模拟研究。

表 1.1 RAFM 钢各成分比例 (%)

	9Cr-2WVTa	F82H	EUROFER97	CLAM
Cr	8.90	7.46	8.82	9.0±0.1
C	0.11	0.09	0.10	0.10±0.02
Mn	0.44	0.21	0.37	0.45±0.05
P	—	—	<0.005	0.003
S	—	—	0.003	0.002
B	—	—	<0.001	—
N	0.021	0.006	0.021	0.02
W	2.01	1.96	1.10	1.5±0.1
Ta	0.06	0.023	0.068	0.07±0.03
Si	0.21	0.10	0.005	0.01
Ti	—	—	0.006	<0.006
V	0.23	0.15	0.19	0.20±0.02
Ni	<0.01	—	0.021	0.02
Co	—	—	0.005	<0.005
Cu	—	—	0.0038	<0.005
Nb	<0.01	0.0001	<0.001	<0.001
O	—	—	0.0026	<0.0026
Mo	0.01	0.003	0.0012	—

1.2 辐照氦效应多尺度模拟方法

反应堆结构材料在受到高能中子辐照时会发生嬗变反应(n, α)^[6], 产生大量的氦, 氦原子不易溶于晶格, 会在缺陷、位错和晶界等位置聚集形成氦团簇(氦泡)。同时, 中子辐照会引起材料内部原子的位移, 导致辐射损伤, 产生大量的缺陷, 这些缺陷与氦泡之间协同的辐照氦效应会导致材料的力学性能发生降级, 加速降低材料的服役寿命^[7]。辐照氦效应对材料性能的主要影响有: 氦致硬化、氦致脆化和氦诱导膨胀^[8], 该过程具有较长的时间尺度(皮秒~几十年)和较大的空间尺度(原子~几米), 对实验模拟的需求较高。因此, 常用多尺度模拟方法^[9]对反应堆结构材料在中子辐照下受到的影响过程进行模拟预测。如图 1.1 所示, 模拟方法有从原子尺度出发的属于第一性原理(First Principle)的从头算分子动力学(AMD)方法^[10], 和模拟原子体系的分子动力学(MD)方法^[11]、分子静力学(MS)方法, 以及基于统计模拟的动力学蒙特卡罗(KMC)方法^[12]和更加宏观的连续介质力学(Continuum Mechanics)、断裂力学(Fracture Mechanics)模拟等。本文旨在研究微观结构和宏观性能之间的关联性, 分子动力学通过严格计算每个原子的牛顿运动方程来模拟平衡和非平衡原子体系的运动, 从原子出发得到体系的性质, 与本文研究内容契合, 因此本文以分子动力学作为基础的研究方法并结合时下热点的深度学习学习方法, 进一步提升了模拟研究的效率。

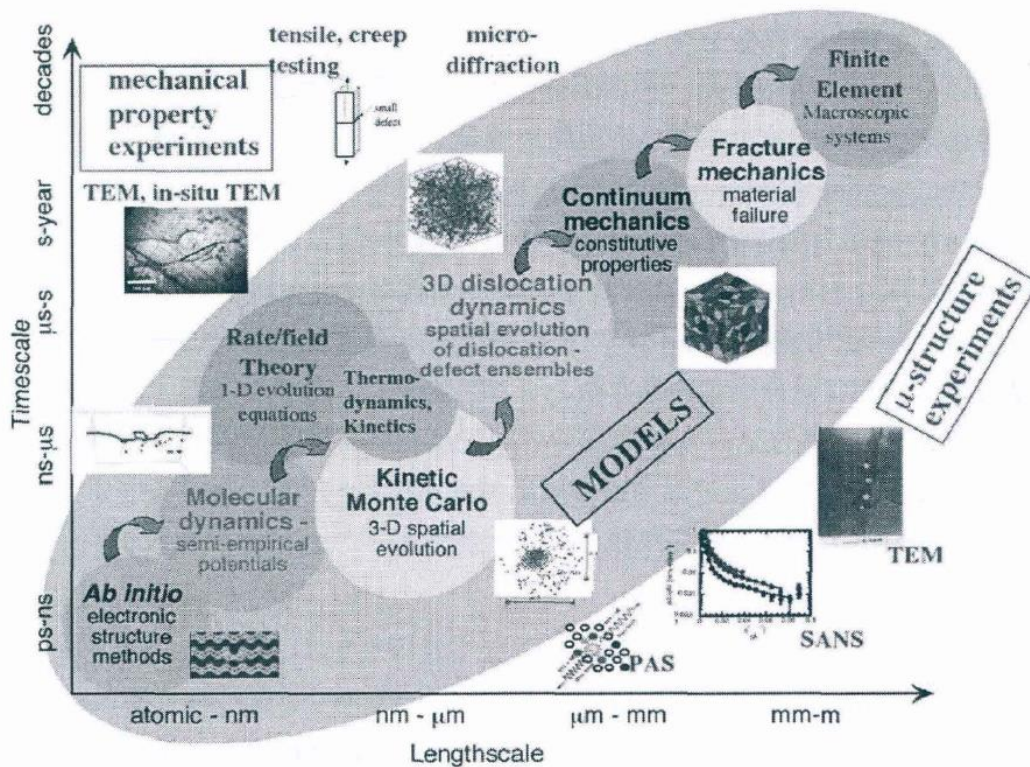


图 1.1 核结构钢辐照效应多尺度模拟方法^[9]

1.2.1 分子动力学方法

分子动力学通过牛顿经典力学方法求解整个体系中各原子的状态,并经过统计办法得到该体系的宏观性质和微观分布^[11]。因此,分子动力学可以得到分子在各种不同的条件下的分布情况,比如在固体表面、晶体晶界和电磁场等环境下的分布和取向。同时,分子动力学还能依据各原子状态的统计信息得到各种体系的热力学性质,比如势能、动能和焓等^[11]。除此之外,分子动力学还可以为体系的演化过程设置一定的约束条件,这些约束条件称为系综(ensemble),常用的有正则系综(NVT)、微正则系综(NVE)、等温等压系综(NPT)和等压等焓(NPH)等^[11]。通过分子动力学可以直接观察单个原子的运动情况,这是实验的方法所不具备的。分子动力学更有助于解释诸如表面、固液界面和晶界等边界处的原子行为,因此,研究者将其作为研究微观原子状态和宏观材料性能间的重要工具,例如, D. Y. Sun, M. Asta 等^[13]基于分子动力学的研究结果提出了计算固液界面能的毛细波动法; Wolf^[14]和 Nakashima、Takeuchi^[15]研究了 Mo 和 bcc-Fe 的对称倾斜晶界的晶界能及其结构之间的关联,并在特定的倾斜角度下观察到了能量峰值; T. Nagano 和 M. Enomoto^[16]计算了 α 铁和 γ 铁的界面能,发现了在特定取向(Kurdjumov–Sachs, K-S 取向和 Nishiyama–Wassermann, N-W 取向)处存在能量峰值。随着计算设备的进步,分子动力学模拟得以在更多的方面实现更为广泛而深入的研究。

1.3 对称倾斜晶界/晶界能

反应堆结构材料通常为多晶体,其力学性能不仅与晶粒尺寸直接相关,而且还取决于晶界的数量和种类。晶界是多晶材料中相邻两晶粒之间的过渡界面^[17],不同于规则排列的晶粒内部原子,晶界原子排列紊乱,容易聚集杂质、位错和点缺陷等。因此,RAFM 钢的晶界处容易聚集氢原子,形成的氢泡对材料的强度和韧性都将造成损害。不同的晶界聚集氢原子的能力不同,研究晶界的微观结构和宏观性质间的关联性能为设计更优性能的结构材料提供帮助。

1.3.1 对称倾斜晶界

晶界具有九个空间自由度,其中四个微观参数决定晶界的能量,剩余五个宏观参数确定晶界的相对位置。在立方晶系中,晶粒取向的运动都是以转动的形式进行的,在一定的旋转轴 $\langle u \ v \ w \rangle$ 上旋转一定的角度 θ 包含了三个自由度,剩下的两个自由度确定晶界平面的取向 $(h \ k \ l)$,通常为该平面的法线方向。当旋转轴 $\langle u \ v \ w \rangle$ 位于晶界平面 $(h \ k \ l)$ 内时,该晶界被称为倾斜晶界,由一系列的刃位错构成。旋转轴 $\langle u \ v \ w \rangle$ 垂直于晶界平面 $(h \ k \ l)$ 的晶界被称为扭转晶界^[18],主要包含了螺位错。同时具有刃位错和螺位错的晶界被称为混合晶界。如果两个晶粒之间的

平面具有相同的取向并按图 1.2 所示互相倾斜 $\theta/2$ ，则称形成的界面为对称倾斜晶界(Symmetric tilt grain boundaries, STGBs)，倾斜角 θ 与晶面指数(h k l)和旋转轴 $\langle u \ v \ w \rangle$ 之间有如下公式：

$$\theta = 2\cos^{-1}\left(\frac{|vh-uk|}{\sqrt{(u^2+v^2)(h^2+k^2+l^2)}}\right) \quad (1.1)$$

其中晶面指数(h k l)和旋转轴 $\langle u \ v \ w \rangle$ 需满足：

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} \cdot \begin{pmatrix} h \\ k \\ l \end{pmatrix} = 0 \quad (1.2)$$

描述晶界结构的模型有重合位置点阵模型(Coincidence Site Lattice, CSL)，O 点阵模型(O-lattice)等^[18]。本文主要以 CSL 模型为依据构建对称倾斜晶界，CSL 模型认为双晶体是通过两晶粒界面运动和不同晶粒内原子的彼此抑制得到的。如图 1.2 所示，在一些特殊的晶界运动中两个晶体的晶格节点是重合的，重合位置点阵便是两个晶体晶格平移子集的交集部分，以重合晶胞体积和原始晶胞体积之比—重合指数 Σ 作为重合位置点阵的表征参数，其随 θ 角是不连续变化的。

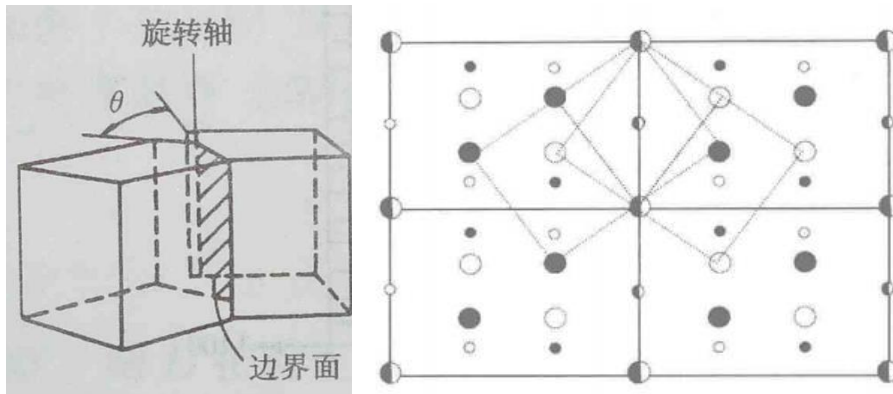


图 1.2 对称倾斜晶界结构^[18]

1.3.2 晶界能

晶界能作为晶界的主要特性之一，可在一定程度上衡量晶界的宏观性质^[19]。晶界能高的晶界往往比晶界能低的晶界要更不稳定，而且高晶界能晶界会通过吸收晶体内的某些杂质原子来降低自身的晶界能，这使得高晶界能晶界更容易引起杂质原子的偏析。因此，“晶界工程”常以晶界能作为晶界性质的衡量参数，通过研究晶界能和晶界微观结构间的关联性达到通过控制晶界结构改变材料性质的目的。在研究两者关联性之前，首先需要对获取晶界能的方法有一个全面的了解。目前，获取晶界能的主要方法有理论推导、实验测量和模拟计算等^[20]三种方法，著名的 Read-Shockley 公式^[21]基于晶界结构进行理论推导得到了晶界能，但仅适用于小角度晶界。Hasson 和 Goux^[22]通过实验测定了 Al $\langle 100 \rangle$ 和 $\langle 110 \rangle$ 对称

倾斜晶界的晶界能随取向差角的变化规律,但该结果受到测量条件的限制,难以大范围运用。因此,若能快速准确地获取晶界能,将为“晶界工程”提供便利,提高实验效率。

随着原子模拟理论的发展,更多地是基于分子动力学来研究晶界微观结构和宏观性质,主要有块状模型^[20]和球形模型^[23],其中块状模型一般具有周期性边界。分子动力学方法所要消耗的计算资源较大,并随晶界体系的增大而增加,难以满足“晶界工程”中计算大量晶界的需求。也有少数研究者^[24,25]尝试从数据科学的角度出发,在分子动力学模拟的基础上构造出合适的特征,并通过机器学习模型^[26]对 Al 的晶界性能进行预测,但其准确性受限于人为构造的特征,优质的特征能带来更好的模型效果。同时,由于模型信息量大,特征复杂,人工提取合适特征的难度较大,较难在大的取向差角范围内得到应用。近年来,深度学习^[27]以更深层的网络结构实现了自动提取特征的能力,成功在人工智能领域取得了非凡的成绩,更有研究者将其应用于生物蛋白质^[28]、超高碳钢^[29]和复合材料^[30]中。因此,本文主要探索了在 α -Fe 中常见的对称倾斜晶界中,通过具有高效提取空间信息能力的深度卷积神经网络算法(CNN)^[27]进行晶界能预测。

1.4 机器学习/深度学习概述

人工智能(Artificial Intelligence)想要创造出能够表现人类相似行为的系统,可以对环境进行观察,然后通过大脑判断后再做出与之对应的行为。机器学习方法^[31]和近年来比较流行的深度学习方法^[27]都为实现真正的人工智能带来了希望。机器学习/深度学习方法基于统计学知识,与多领域专业知识相互结合,推动了各个领域的智能化应用以及本身算法的进步。与经过剖析大量数据而发掘出有用信息的数据挖掘^[32]类似,机器学习/深度学习是建立在大量包含有用信息的样本数据的基础上的。它们需要对数据进行预先处理,去除掉无用的噪声信息,提取出数据的主要特征(特征选择),并对数据的范围进行界定(规范化),通过这些方法提高数据的质量,然后进行数据的分析和挖掘。机器学习/深度学习通过设计高效的算法找到特征空间和输出空间的对应关系,并将其应用在未知的数据中。

传统的机器学习算法主要有以下几种:线性回归模型、Logistic 回归模型、k-近邻算法、决策树、贝叶斯分类器、支持向量机(SVM)、神经网络和集成学习等^[31]。其中,神经网络模仿了人体的生物神经元构造,研究者们在 1943 年提出了著名的 M-P 神经元模型^[33],如图 1.3 所示,该模型通过同时接受输入层传来的多个信号,经过加权求和后与预先设置的阈值进行了比较,在满足一定的阈值后再通过激活函数产生了输出层的信号,并将输出层的信号与预期的目标信号进行比较更新模型的链接权重。1958 年,另有研究者基于该模型的拓扑结构,通过累

叠多个神经元提出了感知机模型^[34]。而单层的感知机难以应对非线性可分的问题，因此研究人员在输出层后又新叠加了一层神经元，与之前的神经元形成了串联，原来的输出层成为了隐藏层(Hidden Layer)，从而得到了两层感知机模型。重复该操作可以得到多层的感知机模型，多层感知机对非线性问题的处理能力远高于单层的感知机模型。

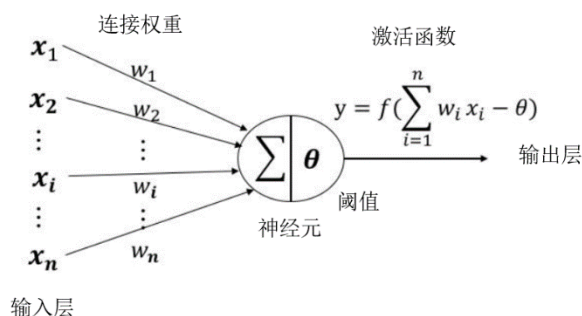


图 1.3 M-P 神经元模型^[33]

深度学习则是在多层感知机模型上又加深了网络的结构^[27]，但这种加深不是简单的堆叠，而是通过一些特定的网络结构在提高模型的拟合能力的同时保证了模型的高效性。但随着模型参数数量的增加，模型的训练出现了问题，比如梯度弥散等问题，梯度无法传递模型的参数就无法更新。因此，研究者在传统的误差逆传播算法^[35]基础上提出了 ReLU 激活函数提高了收敛速度并解决了梯度消失的问题。常见的深度神经网络结构有：卷积神经网络(CNN)^[36]、深度信念网络(DBN)、循环神经网络(RNN)^[37]、递归神经网络(RNN)、生成对抗网络(GAN)等。它们有的被用于目标检测，有的被用于广告推荐，在生活中的各个方面都能看到它们的身影，自然语言处理中的 Transformer 模型更是为语义理解等领域带来了巨大的发展。

深度学习方法相比于传统的机器学习方法，可以分析更大规模的数据，不再强调数据的先验特性，试图从更基础简单的数据中自动地进行特征学习，提取出更适合具体任务的全新特征，提高了学习的效率。

1.5 本文研究意义和主要内容

本文的研究立足于晶界工程在控制 RAFM 钢辐照氦效应上的应用，主要围绕着 RAFM 钢中的晶界结构展开，结合新兴的深度学习算法，探索更加高效地研究晶界性能和材料辐照效应的途径，为聚变堆材料中的晶界工程应用研究添砖加瓦。本文主要的研究内容有两块：1.基于 TextCNN 预测 α -Fe 中对称倾斜晶界的晶界能；2.基于 LSTM 判断间隙氦原子在晶界处的迁移状态。通过预测晶界能

的研究可以探索晶界微观结构与晶界宏观性能之间的关联性,利用具有高效提取空间信息能力的卷积神经网络,提取出更加简单有效的晶界结构特征,为进一步应用在晶界工程中提供方案。间隙氦原子迁移状态的研究主要是通过具有处理时间序列能力的循环神经网络来判断氦原子在晶界中经过一段时间的状态改变后是否发生迁移,并通过该任务来分析影响间隙氦在晶界中的迁移状态的主要因素。为辐照氦效应的多尺度模拟研究带来了全新的方法,在加速分子动力学领域有着广阔的发展空间。

全文共有 5 个章节,第 1 章节“绪论”重点介绍了研究材料:RAFM 钢,基础研究方法:辐照氦效应多尺度模拟方法,主要研究对象:对称倾斜晶界/晶界能,提升研究方法:机器学习/深度学习方法。其中,基础研究方法中的分子动力学作为后续研究的样本来源被进行了详细的介绍,在研究对象中主要介绍了对称倾斜晶界和获取晶界能的方法。最后阐明了本文的研究意义和研究内容。第 2 章节“机器学习/深度学习原理”对支持向量机、卷积神经网络、循环神经网络等算法分别进行了原理的介绍。首先介绍了整个方法的流程,再构建了倾斜轴为 $\langle 100 \rangle$ 和 $\langle 110 \rangle$ 的 92 个 α -Fe 对称倾斜晶界,并以晶界结构为基础构造了三种晶界特征参数(径向分布函数、中心对称参数和积累中心对称参数),然后详细介绍了 TextCNN 模型的原理和具体搭建过程,提出了按倾斜角分层抽样和数据增强等提升模型性能的方法。最后对该任务的结果进行了分析与讨论,并通过与传统机器学习方法的对比重点分析了各晶界特征参数的优劣。第 4 章节“基于长短时记忆网络研究间隙氦在晶界中的迁移”通过介绍循环神经网络和间隙氦迁移机制实现了判断间隙氦是否迁移的模型,并通过该模型在各种晶界环境中的分类情况研究了决定间隙氦在晶界中是否迁移的主要因素。构建了间隙氦原子在完整 α -Fe 晶粒、 $\langle 110 \rangle$ 倾斜轴上的 $\Sigma 3 \{112\}$ 晶界和 $\Sigma 11 \{113\}$ 晶界中演化 1ns 的迁移模型。在第 5 章节“总结与展望”中重点描述了下一阶段的研究方向。

第 2 章 机器学习/深度学习方法原理

在这一章节中, 本文对主要的三种机器学习/深度学习方法的原理进行了逐一的介绍。

2.1 支持向量机(SVM)

支持向量机模型^[58]通过支持向量可以找到能区分类别的超平面, 并能最大化类间距离^[58], 具有更好的适应性和更高的分类精度, 保证最小的分类错误率。由于 SVM 找到的是分类界面, 而本文的目标是预测晶界能, 属于回归模型的范畴, 因此本文选用了支持向量回归(SVR)模型。下面从分别对 SVM 和 SVR 的基本原理进行描述。

如图 2.1 所示, SVM 的训练样本分为两类假设为 $y_i \in \{-1, +1\}$, 各样本输入特征为 $x_i \in R^n$, 其中 n 为输入特征的维度。SVM 期望的超平面为:

$$wx_i + b = 0 \quad (2.1)$$

能使得所有样本满足:

$$wx_i + b \geq 0, y_i = +1 \quad (2.2)$$

$$wx_i + b < 0, y_i = -1 \quad (2.3)$$

同时, 该超平面还希望能与各类别的样本保持最远的距离, 两类样本间的最小距离被称为“间隔”, 具体公式为:

$$\gamma = \frac{2}{\|w\|} \quad (2.4)$$

通过最大化间隔能够找到该超平面:

$$\max \frac{2}{\|w\|} \quad (2.5)$$

$$\text{s.t. } y_i(wx_i + b) \geq 1, i = 1, 2, \dots$$

具体的求解过程涉及到该问题的对偶问题、拉格朗日乘子法和序列最小优化(SMO)算法等, 这里不做详述。通过核函数和设置软间隔还可以使 SVM 处理非线性和不可分问题。

与 SVM 不同的是, SVR 中的样本都属于同一类。通过预先对超平面设置一定的偏差 ε 形成一个间隔带, 并期望间隔带能尽可能地宽, 这样更多的样本能落到该区间内, 对在该间隔带里的样本不计算它们的损失, 而对该区间外的样本则根据它们到间隔带的距离计算损失。具体的损失函数如下:

$$\text{loss} = \begin{cases} 0, & \text{if } |z| \leq \varepsilon \\ |z| - \varepsilon, & \text{otherwise} \end{cases} \quad (2.6)$$

其中 z 为间隔带外样本到超平面的距离。因此, SVR 的损失函数为:

$$\min \frac{1}{2} ||w||^2 + \text{loss} \quad (2.7)$$

实际任务中, 往往很难确定一个合适的偏差 ε , 因此常通过加入松弛变量使得间隔要求被放松。具体的求解过程和 SVM 相类似, 这里也不做详述。

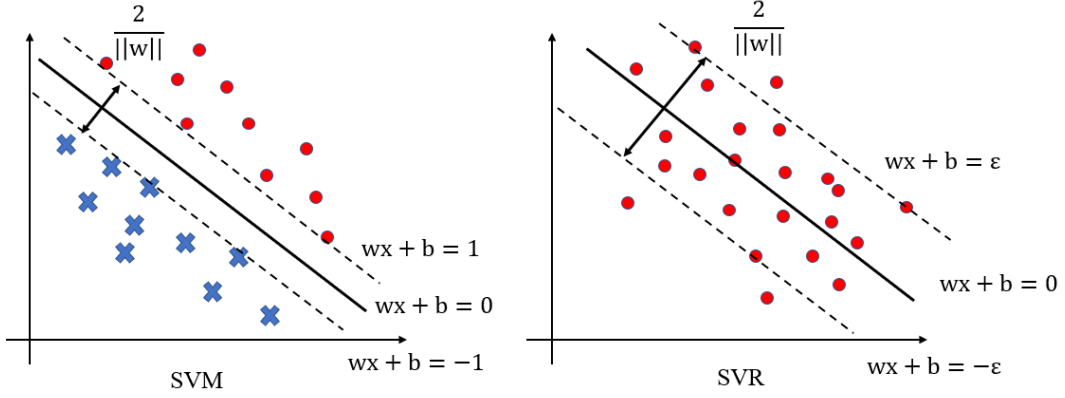
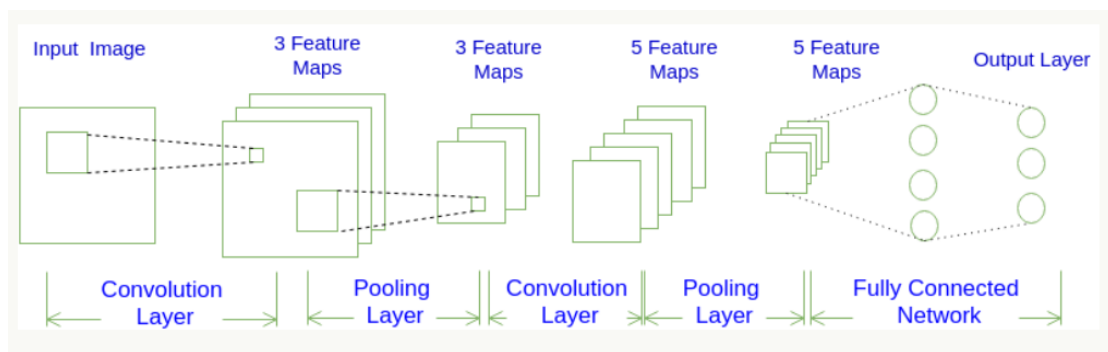


图 2.1 SVM 和 SVR 原理示意图

2.2 卷积神经网络(CNN)

2.2.1 模型基本原理

构建完晶界特征后, 本文挑选了适合处理空间信息的卷积神经网络(CNN)作为预测晶界能的基础模型。卷积神经网络在经历了早期的低谷期后于 2012 年被研究者们进行了深度的研究并提出了多种相关的模型结构, 比如谷歌的 GoogleNet^[45]、微软的 ResNet^[46]等, 围棋软件 AlphaGo^[47]也用到了这种网络。卷积神经网络善于处理信号间具有空间相关性的数据, 前文构建的三种晶界特征均有特定的空间相关性, 中心对称参数和积累中心对称参数是关于 y 轴坐标的相关性, 而径向分布函数则是关于径向距离的相关性, 因此利用该网络能够很好得提取它们所包含的空间信息。卷积神经网络利用卷积核连接了局部输入^[36], 同一卷积核在整个输入上滑动计算卷积, 相比于传统的神经网络降低了模型的参数量, 并且卷积神经网络还会考虑输入特征间的位置信息。如图 2.2 所示, 基础的卷积神经网络由卷积层(Convolution)、激活层(Activation)、池化层(Pooling)、全连接神经网络(FCN)等^[27]结构组成。

图 2.2 卷积神经网络示意图^[27]

卷积层通过卷积运算来替代线性相加，二维离散的卷积运算有：

$$O(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (2.8)$$

其中， O 是卷积后输出的特征映射值， I 为输入的二维网格数据， K 为卷积核函数， i 、 j 、 m 、 n 分别是网格坐标和核函数格点坐标。卷积核尺寸往往比输入数据更小一些，输出单元仅与部分输入单元关联，实现了稀疏交互的特征^[27]，并且所有的输入单元均通过该卷积核进行输出，同一卷积核的参数在整个输入数据上作用，实现了参数共享的特点。具体的三通道的二维卷积层在三通道图像 ($3 \times 3 \times 7$) 上进行卷积计算的过程是：设置 2 个 ($3 \times 3 \times 3$) 大小的卷积核，在图像上移动，不断地与被卷积核包含的元素进行卷积计算，最终将各通道结果叠加起来。

卷积后得到的特征映射需要通过激活层的激活函数引入非线性，常见的激活函数有修正线性单元 (ReLU, Rectified Linear Unit)^[48]、Sigmoid 函数、tanh 函数^[49] 等。ReLU 的定义为 $f(x) = \max(x, 0)$ ，可见它在小于零的区间为 0，并且导数也为 0，而在大于 0 的区间，导数为 1。这使得 ReLU 作为激活函数在计算导数时所需的计算代价相对于别的激活函数来说要小得多。同时，在模型利用反向传播算法进行梯度计算时，ReLU 特殊的导数能够减轻梯度消失和梯度弥散等现象。

池化层作为一种下采样操作，主要用于降维，使用某一局部位置的总体特征来替代原位置的特征，大大压缩了参数量，提高了网络预算速度。常见的池化操作有主要最大值池化 (Max Pooling)^[50]、平均值池化 (Average Pooling)^[51] 等。最大值池化就是只保留一定区间内的最大值，按这个思路进行扩展，还可以选取区域内最大的 K 个值，称为 K-Max-Pooling^[50]。K-Max-Pooling 在保留最大 K 个值的同时还会保留这些特征值原始的先后顺序，供后续阶段使用。Chunk-Max-Pooling 是通过把每个卷积核输出的特征向量进行区间划分，而后分别在这些区间里各取一个最大的特征值，如图 2.3 所示，将输出的 6 维特征向量划分为三段 (图中各颜色代表不同的分段)，各段取最大的一个，按序组成三维特征向量。在很多任务中，池化层还承担了改变参数个数的作用。也有一些研究^[51] 指出了在各种情况下

应用哪类池化函数更优。

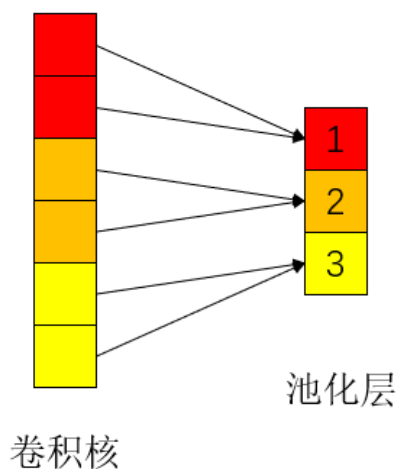


图 2.3 Chunk-Max-Pooling 示意图

最后将卷积神经网络输出的特征展开后输入全连接神经网络，用全连接层拟合输入样本特征和输出结果，实现最后的分类或回归。在本文的研究中，最终需要的是预测晶界能，各晶界能间具有大小差异且范围较广，不利于分成多个类别进行预测，因此通过全连接层直接输出了晶界能大小。若应用在分类问题上，可以输出各类别的概率并在各类别概率间进行归一化操作，常见的有 Softmax 函数^[52]。

2.2.2 模型训练过程

在确定完整个 CNN 模型的结构后，还需要为模型的训练过程设定一些超参数，包括批大小(Batch Size)、迭代次数(Epoch)、学习率(Learning Rate)、优化器(Optimizer)和损失函数(Loss Function)^[53]等。整个模型的训练过程如图 2.4 所示，将训练样本输入网络中得到模型的预测输出，并以损失函数计算它与目标输出之间的损失(loss)，loss 越大代表模型预测得越不准确。最后，以 loss 作为评判依据，并通过优化器去更新模型的参数，使得模型预测结果和预计输出能更加接近，降低两者间的 loss。这是一个不断学习的过程，模型参数在每一次训练中得到更新，最终达到最优的数值。这需要大量样本作为支撑，因此往往将现有的样本重复循环地输入网络，迭代次数则代表了重复的次数。在本次实验中，epoch 设置为 1000，既保证了模型参数学习已经收敛同时也防止了多度训练可能造成的过拟合现象。

实际模型训练中，往往将多个样本合并成一个张量矩阵(批处理，Batch

Training)进行计算,这样能够提高模型的计算效率,充分利用有限的计算机内存。同时,批处理是通过一个批次上的平均效应来对参数进行更新的,在一定程度上能够减小模型训练时形成的震荡。当数据较大时,还可以通过批归一化(BN)^[54]改善数据分布,降低不稳定性,训练出更有效的模型,由于本文的数据量较小,则未设置批归一化层,批大小设置为 10。

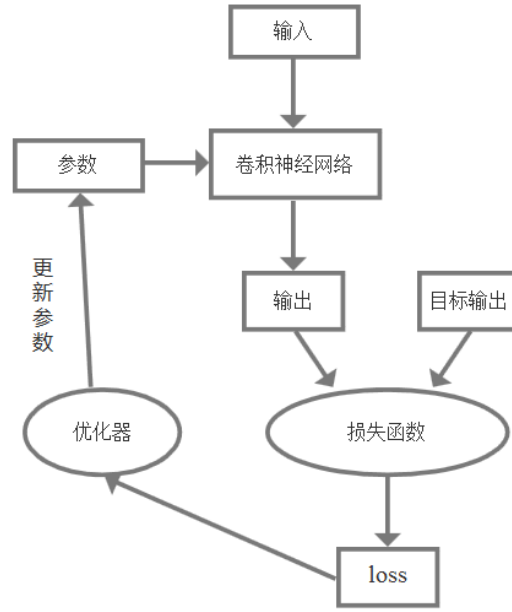


图 2.4 卷积神经网络模型训练流程图

损失函数用来度量模型预测结果和预计输出之间的差距。随着模型训练程度的加深,预测结果逐渐靠近预期输出,损失(loss)也逐渐减小并收敛到一个极小值,此时的模型参数即为训练的最终结果,因此网络训练的目标是通过调整参数最小化损失函数 loss。回归模型常见的损失函数^[27]有平均绝对值误差(MAE, 又称 L1 损失)、均方误差(MSE, 又称 L2 损失)、平滑平均绝对误差(Smooth MAE, 又称 Smooth L1 损失)和 Log-Cosh 损失等。L1 损失相对于 L2 损失,对异常值表现得更加稳健,但由于 L1 在靠近 0 值时仍有较大的梯度,则可能在参数更新时丢失最小值,而 L2 随着 loss 接近其最小值,梯度逐渐减小,更利于找到模型的最优参数。因此,本文选择了结合两者长处的平滑 L1 损失,其计算公式为:

$$\text{loss}(x, y) = \frac{1}{n} \sum_{i=1}^n \begin{cases} 0.5 * (y_i - f(x_i))^2, & \text{if } |y_i - f(x_i)| < \delta \\ \delta |y_i - f(x_i)| - 0.5\delta^2, & \text{otherwise} \end{cases} \quad (2.9)$$

其中, n 为样本个数, x 为样本输入特征, y 为预期目标输出, $f(x_i)$ 为模型预测结果,通过调整超参数 δ 可以改变误差为均方误差的范围,这里设置为 10。

优化器的目的是找到能让 loss 缩小的参数调整方法,它会根据 loss 和网络的

情况,更新网络的参数,并让 loss 往更小的值去靠近。最开始是通过随机调整模型参数,然后重新计算 loss,当 loss 减小时则保留这个参数。但该方法的效率是十分低的。因此研究者提出了梯度下降的更新参数方法,通过计算损失函数在各参数上的偏导数来确定应该往哪个方向调整模型参数,在每一步位置上选择最易下降的方向,并最终到达最低点。实际训练时,每次下降都通过计算整个训练集的偏导数需要耗费的计算资源较大,因此往往随机挑选一定数量的样本,在这些样本上计算梯度下降的平均效应,这被称为随机梯度下降(SGD)^[27]。

梯度下降的具体公式为:

$$\omega_{t+1} = \omega_t - \eta \frac{\partial \text{loss}}{\partial \omega_t} \quad (2.10)$$

其中, ω_t 为更新前的参数, ω_{t+1} 为更新后参数, η 为学习率(Learning Rate)。调整学习率可以改变每一次下降的幅度,学习率越大每次下降的幅度越大。过大的学习率可以会导致错过 loss 最低点,而过小的学习率则需要耗费较大的计算代价,因此找到两者间的平衡点是十分重要的。在 Pytorch 中,可以通过 MultiStepLR 模块来设置阶梯状的学习率,本文通过该方法对于中心对称参数特征将前 100 个周期的学习率设置为 $1e^{-4}$,之后设置为 $1e^{-5}$,而对于积累中心对称参数特征学习率设置为 $3e^{-4}$,径向分布函数的学习率设置为 $1e^{-5}$ 。

从数学角度而言,梯度下降算法是一种贪心算法^[31],对于一些非凸(non-convex)问题,有可能陷入局部极小值点。初始点设置位置不同则可能收敛到两个不同的局部极小值点,要想获得全局极小值点最直接的方法则是设置多个初始位置,再从所有结果中挑选出全局极小值点。但这种方式耗时耗力,而且也并不能保证最终的结果就是全局极小值。因此研究者发展了一系列在随机梯度下降基础上改进的算法,比如 AdaGrad^[55]、RMSProp、Adam^[56]等。其中 AdaGrad 是 John Duchi 等于 2011 年提出的,它不再对所有参数设置一样的学习率,而是根据参数改变的情况为其分配更适合的学习率,为频繁变化的参数分配更小的学习率而为稀疏的参数分配更大的学习率^[55]。RMSProp 依据梯度平方的指数移动平均数来调整学习率^[56],这种方法不会出现 AdaGrad 中梯度迅速减小的情况,在不稳定的目标函数下有更好的调节学习率的能力。Adam 综合了这两种算法的优点,在多数情况下都表现得非常优异,而且对内存的需求也更低。Adam 算法的伪代码如下:

Adam 算法

输入: α (步长)

β_1 、 β_2 (矩估计的指数衰减速率,在 $[0, 1)$ 之间)

$f(\theta)$ (带有参数 θ 的随机目标函数)

θ_0 (初始参数向量)

$$m_0 \leftarrow 0 \text{(初始一阶矩向量)}$$

$$v_0 \leftarrow 0 \text{(初始二阶矩向量)}$$

$$t \leftarrow 0 \text{(初始时间步)}$$

过程:

while θ_t 没有收敛 do

$$t \leftarrow t + 1$$

$$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1}) \text{ \# 通过随机挑选目标计算梯度}$$

$$m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t \text{ \# 更新一阶矩估计}$$

$$v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \text{ \# 更新二阶原始矩估计}$$

$$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t) \text{ \# 计算偏差修正过的一阶矩估计}$$

$$\hat{v}_t \leftarrow v_t / (1 - \beta_2^t) \text{ \# 计算偏差修正过的二阶原始矩估计}$$

$$\theta_t \leftarrow \theta_{t-1} - \alpha \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon) \text{ \# 更新参数}$$

end while

return θ_t (最终结果模型参数)

从上面的伪代码可以看到, Adam 算法同时计算了梯度的均值(一阶矩估计)和梯度的方差(二阶矩估计)^[56], 在每一个更新参数过程中, 通过当前的梯度更新这两个参数, 再依据这两值的修正项共同更新模型参数。值得注意的是, 并没有绝对更好的优化算法, 还是应该依据实际任务挑选更适合的优化器, 这样才能最大程度的发挥各种算法的优势。本文根据实际需求选择了 Adam 优化器。

2.3 循环神经网络

循环神经网络(RNN) 与之前介绍的卷积神经网络一样也是深度学习中的重要模型。它的出现是为了解决输入信号间具有先后顺序的问题, 比如自然语言、

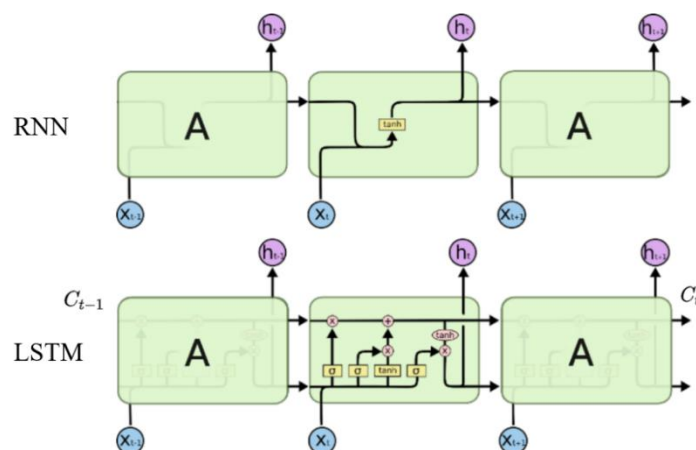


图 2.5 RNN 和 LSTM 模型基本结构单元^[37]

时间序列等。循环神经网络会在同一层的神经元间传递信息^[37],具有存储信息的能力,而传统的神经网络只能在层与层之间单元进行连接,同层的信号不具有先后关系。RNN 通过连续性得输入有关联性的特征序列获得了处理时间序列数据的能力,其基本结构单元如上图 2.5 所示。

图中 A 代表 RNN 模型结构, \tanh 为模型的激活函数, X_t 表示模型在该时刻的输入特征, h_t 表示模型在该时刻的输出特征,而该输出特征同时与下一时刻的 RNN 模型输入端相连,前一时刻的特征与该时刻的输入特征一起输入到模型中进行学习,通过这样的“内环”保留了历史时刻的输入信息,并且在当前时刻的网络中得到了应用,这样独特的设计提高了模型的能力。但在实际应用的过程中,长时间的循环使得上下文的信息更偏重于距离较近的输入特征,而较远的特征则会由于时间间隔太长而被丢失。因此,研究者^[65]在原始模型的基础上改进了重复的 RNN 结构,并将其命名为长短时记忆网络(LSTM),它能够记忆更早时刻的信息。如图 2.5 所示, LSTM 将循环神经网络模型中单一的层结构(\tanh 层)扩展为三个“门”结构:遗忘门、输入门和输出门^[37],并通过细胞状态(cell state, C_t)来保留上下文的信息。每个门结构都通过一个 sigmoid 单元来描述信号的通过量,其中遗忘门是对历史信息 C_{t-1} 进行了筛减,输入门则决定了哪些新输入的信息将加入到当前细胞状态 C_t ,而输出门则计算了当前网络的输出。通过这三种“门”结构, LSTM 成功得记住了长期的信息。

LSTM 的训练过程和卷积神经网络过程相似,这里就不作重复的介绍。

2.4 本章小结

在本章中对机器学习方法中的 SVM 和深度学习方法中 CNN 和 LSTM 的原理进行了详细的描述,同时对深度神经网络的训练过程进行了说明。

第 3 章 基于卷积神经网络研究晶界能和晶界结构相关性

本章主要介绍了通过深度学习方法预测 α -Fe 对称倾斜晶界晶界能的途径。下文首先对整个流程进行一个系统的介绍，提出主要的技术特点，然后从提取晶界特征、设计 TextCNN 网络这两个方面进行阐述，最后在测试集上对预测模型结果进行了评价，分析了按倾斜角分层抽样和数据增强方法对模型的影响，并与传统的机器学习方法进行了对比。

预测晶界能的深度学习模型流程如图 3.1 所示，主要包含三个阶段：第一阶段，依据对称倾斜晶界的晶界参数在分子动力学模拟中构建重合位置点阵模型，重复进行以建立起对称倾斜晶界库；第二阶段，根据分子动力学模拟中得到的对称倾斜晶界的原子坐标计算中心对称参数，之后融合进原子密度信息构造出晶界特征参数如积累中心对称参数等；第三阶段，将完成特征构建的样本库按倾斜角进行分层抽样划分为训练集和测试集，并通过构造 TextCNN 模型在数据增强后的训练集上进行学习、验证，最终以测试集上预测的晶界能和分子动力学模拟得到的晶界能间的相对差值来评估模型的真实预测效果。下面依次对这三个阶段进行介绍。

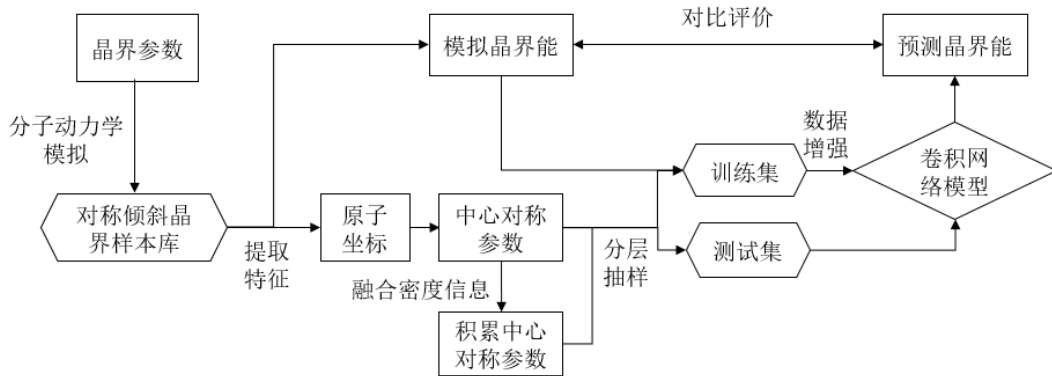


图 3.1 预测晶界能流程图

3.1 建立 α -Fe 对称倾斜晶界库

本文通过 LAMMPS 软件^[38]生成稳定的对称倾斜晶界结构，并得到相应的晶界能。根据重合位置点阵模型，对称倾斜晶界的主要参数有倾斜角 θ 、晶面指数 (hkl) 、旋转轴 $\langle uvw \rangle$ 和重合指数 Σ 。其中，本文选取了旋转轴 $\langle 100 \rangle$ 和 $\langle 110 \rangle$ 的晶界，随后通过 python 脚本自动搜取出所有重合指数 Σ 小于 200 的晶面指数，

伪代码如下：

计算晶界参数伪代码

Input: 旋转轴 $\langle u \ v \ w \rangle$ # $\langle u \ v \ w \rangle = \langle 110 \rangle$ 或者 $\langle 100 \rangle$

Output: 晶面指数 $(h \ k \ l)$ # 其中 $\langle 110 \rangle$ 晶面指数有 $(h = k)$, $\langle 100 \rangle$ 晶面指数有 $(h = 0)$

Output: 重合指数 Σ

$\Sigma = 0$ # 初始化重合指数

$h = k = 1$ # 初始化晶面指数，这里以 $\langle 110 \rangle$ 倾斜轴为例

$l = 1$ # 初始化晶面指数

while $\Sigma < 200$: # 这里搜取的是重合指数小于 200 的晶界

 if Σ 为奇数: # 判断 Σ 奇偶

 for 所有 k 在 $(1, \Sigma/2)$ 内 do:

 for 所有 l 在 $(1, \Sigma/2)$ 内 do:

 if 满足 $(u^2 + v^2 + w^2) \cdot (k^2) + l^2 = \Sigma$ 或者 $u^2 + v^2 + w^2 \cdot (k^2) + l^2 = 2 \cdot \Sigma$:

 return k, l, Σ

 end for

 end for

$\Sigma \leftarrow \Sigma + 1$

end while

随后通过公式 3.1 得到倾斜角 θ :

$$\theta = 2 \tan^{-1} \left(\frac{k}{l} \sqrt{u^2 + v^2 + w^2} \right) \quad (3.1)$$

最后依据晶面指数 $(h \ k \ l)$ 和旋转轴 $\langle u \ v \ w \rangle$ 得到了晶界周期方向，三者互相垂直。按已得的基本晶界参数在 LAMMPS 中生成了两个相互对称的 α -Fe 对称倾斜晶粒，并按图 3.2 在模拟空间中进行拼接得到块状双晶模型：

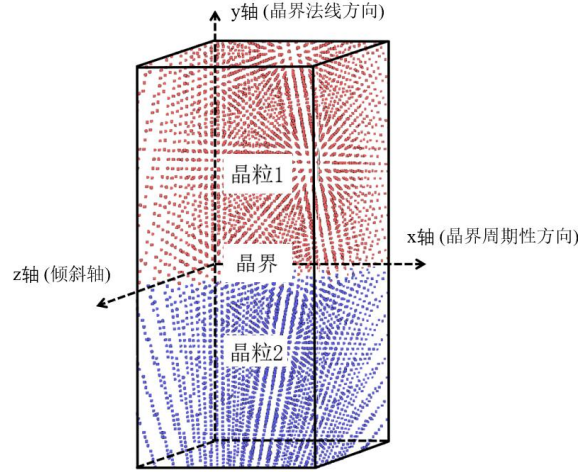


图 3.2 LAMMPS 晶界结构

上图中晶粒 1 和晶粒 2 的交界面即为晶界所在的位置，由于周期性结构，模块的上下两侧也各自存在半个晶界，但这样拼接得到的晶界仍不是稳态的晶界，需要采取一定措施才能使晶界得到稳态的构型。本文依据晶界平面内平移(In-plane Translations) 和原子删除判据(Atom Deletion Criteria) 理论^[39]，使上图中的晶粒 1 相对于晶粒 2 在 z 轴和 x 轴方向进行了刚性平移，平移矢量的大小依据完整花样移动(DSC)点阵^[17]，保证了边界原子位置保持不变。经过相对刚性平移后的两晶粒原子会出现重合度较高的情况，这与实际情况不符，因此原子删除判据通过设定一系列的删除标准，将两个原子间距离小于该标准的两原子随机删除一个，以此枚举出大量可能的构型。最后，在 0K 下对所有构型进行了静力学弛豫并微调了晶界周边原子的位置，通过计算晶界结构和具有相同原子个数的完整晶粒间的势能差得到晶界能：

$$\gamma_{gb} = \frac{E_{gb} - (E_1 + E_2)}{2 * A} \quad (3.2)$$

其中 γ_{gb} 代表晶界能， E_{gb} 是含对称倾斜晶界双晶系统的势能， E_1 和 E_2 分别是上下两个晶粒单独存在时的势能， A 为晶界平面的大小。由于周期性边界的存在，为避免相邻晶界间存在互相影响，遵循两晶界在晶界法向上间距大于 12 nm 的原则。其中铁原子间相互作用以 Ackland 等发展的铁-铁势函数 Ackland04^[40]描述。

如表 3.1 所示，一共构建了 92 个具有最低能量构型的 α -Fe 对称倾斜晶界结构，其中包括倾斜轴 $\langle 100 \rangle$ 的晶界结构 32 个和倾斜轴 $\langle 110 \rangle$ 的晶界结构 60 个，至此组成了 α -Fe 的对称倾斜晶界库。依据表 3.1 中的参数统计了晶界能随倾斜角的分布，如图 3.3 所示。这与 Yasushi SHIBUTA 等^[41]构建的体心立方晶体的对称倾斜晶界的晶界能随倾斜角的分布相接近。图 3.3(a)， α -Fe 中倾斜轴为 $\langle 100 \rangle$ 的对称倾斜晶界的晶界能随倾斜角 θ 的增加呈现一个先上升后下降的趋势，并在

$\theta=48^\circ$ 左右达到顶峰值，与对比文献中最大值出现在 $\theta=45^\circ$ 附近相吻合。同时，在 $\theta=36.87^\circ(\Sigma=5)$ 、 $\theta=53.13^\circ(\Sigma=5)$ 和 $\theta=61.93^\circ(\Sigma=17)$ 处出现了明显的降低情况，反观其余较小 Σ 值的晶界 $\theta=28.07^\circ(\Sigma=17)$ 、 $\theta=67.38^\circ(\Sigma=13)$ 等未出现较小的谷值，不过 Σ 值较小的晶界出现较小谷值的几率更大一些。同样在图 3.3(b)中， $\alpha\text{-Fe}$ 中倾斜轴为 $\langle 100 \rangle$ 的对称倾斜晶界的晶界能随倾斜角 θ 的增加呈现三个峰值的趋势，在 $\theta=50.48^\circ(\Sigma=11)$ 和 $\theta=109.47^\circ(\Sigma=3)$ 处出现了最小值，其中 $\Sigma=3$ 的晶界能为 260mJ/m^2 是整个倾斜角范围内的最低值，究其原因，出现较低晶界能的原因可能与 $\theta=109.47^\circ(\Sigma=3)$ 对称倾斜晶界是孪晶结构有关，这与 D. Wolf^[42]早期研究结果一致。在 $\theta=33.72^\circ(\Sigma=107)$ 附近达到了最大峰值，这与对比文献中最高晶界能出现在 $\theta=140^\circ$ 处不一致，在峰值附近的 $\theta=31.59^\circ(\Sigma=27)$ 出现了凹点，同样的峰值凹点出现在 $\theta=70.53^\circ(\Sigma=3)$ 和 $\theta=141.06^\circ(\Sigma=9)$ 处，可以看到较小 Σ 值的晶界更容易出现相较于近邻倾斜角晶界的晶界能低值。更普遍的情况是，相邻倾斜角的晶界具有更加接近的晶界能，因此本文基于该特点设计了按倾斜角分层抽样的方法，以求得训练样本空间更加得丰富。

表 3.1 $\alpha\text{-Fe}$ 对称倾斜晶界各晶界参数

倾斜角 θ	Σ	晶界能 mJ/m^2	倾斜角 θ	Σ	晶界能 mJ/m^2	倾斜角 θ	Σ	晶界能 mJ/m^2
倾斜轴 $\langle 100 \rangle$ (*为测试集，其余为训练集)								
9.52	145	729	36.87	5	1008	64.01	89	1142
11.42	101	794	41.11	73	1154	67.38	13	1108
12.68	41	825	43.6	29	1181	71.08	37	1058
14.25	65	873	46.4	29	1183	73.74	25	1002
16.26	25	908	47.92	97	1197	75.75	65	945
*18.92	37	972	48.88	73	1192	77.32	41	900
22.62	13	1005	53.13	5	1112	*79.61	61	820
25.06	85	1071	57.22	157	1175	81.2	85	754
28.07	17	1092	*58.11	53	1168	82.37	113	699
31.89	53	1114	59.49	65	1166	83.27	145	653
*33.4	109	1108	61.93	17	1118			
倾斜轴 $\langle 110 \rangle$ (*为测试集，其余为训练集)								
10.1	129	901	58.99	33	1253	114.53	171	729
11.54	99	960	*61.02	97	1280	115.7	113	789
*13.44	73	1027	62.44	67	1282	117.56	67	869
14.65	123	1062	64.3	113	1309	*118.98	97	919
20.05	33	1205	70.53	3	1308	121.01	33	978
26.53	19	1312	75.29	193	1325	124.12	41	1043
28.84	129	1343	77.88	81	1305	126.41	123	1095
*31.59	27	1245	80.63	43	1255	129.54	11	1113
33.72	107	1351	*82.95	57	1262	132.31	153	1180

34.89	89	1343	86.63	17	1207	136	57	1215
36.15	187	1344	89.42	99	1184	141.06	9	1167
38.94	9	1286	90.58	99	1167	145.11	89	1236
*42.18	139	1282	93.37	17	1110	*146.28	107	1232
42.89	187	1272	95.45	179	1062	148.41	27	1217
44	57	1244	97.05	57	1022	*151.16	129	1207
47.69	153	1160	*99.36	43	943	153.47	19	1154
50.48	11	1019	100.96	121	881	159.95	33	1106
*53.59	123	1158	102.12	81	831	166.56	73	962
54.43	153	1178	104.71	193	700	168.46	99	896
55.88	41	1207	109.47	3	260	169.9	129	844

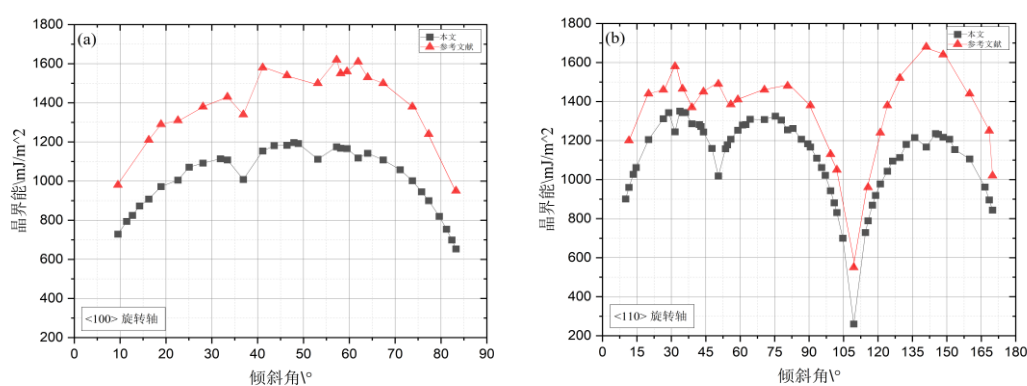


图 3.3 α -Fe 中倾斜轴为(a)<100>和(b)<110>的对称倾斜晶界倾斜角与晶界能关系图^[41]

至此通过分子动力学方法构建了 92 个具有最低能量构型的 α -Fe 对称倾斜晶界结构，其中包括倾斜轴<100>的晶界结构 32 个和倾斜轴<110>的晶界结构 60 个。

3.2 提取晶界结构特征

α -Fe 对称倾斜晶界库主要包含了稳态的晶界结构，由整个双晶模型的原子位置信息和原子特征构成。因为未在晶界中加入杂质，本文仅选择了双晶模型的原子在三维空间中的分布(如图 3.2 中各原子所示)作为基础特征进行下一步研究。晶界平面附近的 Fe 原子分布并不均匀，直接以空间结构进行学习有效性极低，因此，为了提高模型的学习效率，本文尝试将三维的原子坐标信息简化为适合输入的二维数据，并分别构造了三种特征(RDF、CS、ACS)进行对比分析。

3.2.1 径向分布函数

首先，本文通过在晶体材料中常见的径向分布函数作为晶界特征，RDF 又被

称为对关联函数或对分布函数常被用来研究物质的有序性^[17]。如图 3.4 所示，径向分布函数 $g(r)$ 代表选定一个参考原子后，在距离该原子 r 到 $r+dr$ 的球壳中能找到原子的可能性。在该球壳中包含的原子数目 $dn(r)$ 与 $g(r)$ 有下式的关联：

$$dn(r) = \frac{N}{V} g(r) 4\pi r^2 dr \quad (3.3)$$

其中 N 为总的原子数目， V 为模型的体积。

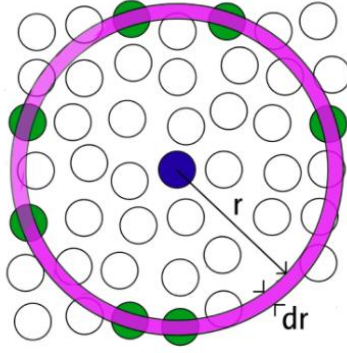


图 3.4 离散空间的径向分布函数示意图^[38]

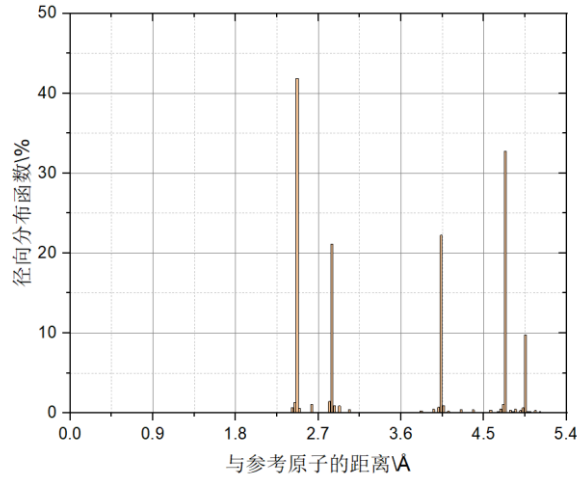


图 3.5 对称倾斜晶界的径向分布函数

实际应用到对称晶界模型，本文通过 Lammmps 在对称倾斜晶界模型的基础上计算了截断距离 (5.4 Å) 内的径向分布函数，并将两晶粒原子同时进行考虑，得到图 3.5 所示结果。

3.2.2 中心对称参数/积累中心对称参数

考虑到块状晶界模型的周期性边界条件，本文通过中心对称参数 (CS, Centrosymmetry Parameter)^[43] 来替代空间原子分布，其计算公式如下：

$$CS = \sum_{i=1}^{N/2} |\mathbf{R}_i + \mathbf{R}_{i+\frac{N}{2}}|^2 \quad (3.4)$$

其中 N 是每个原子的近邻原子个数，体心立方晶体固定为 8， \mathbf{R}_i 、 $\mathbf{R}_{i+\frac{N}{2}}$ 是中心原子到一对处于对称位置的近邻原子的矢量对。中心对称参数可用来衡量该原子周边环境的紊乱程度，如图 3.6 中蓝色点所示，在晶界处有较大的正值，晶格内部原子一般接近于 0，由于热扰动的存在，一般不会直接为 0。

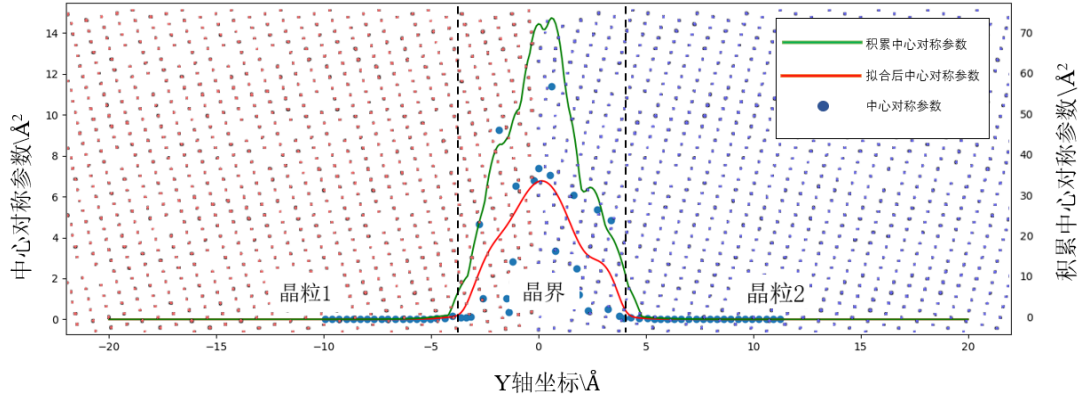


图 3.6 $\langle 110 \rangle$ 倾斜轴的 $\Sigma 9$ 对称倾斜晶界上中心对称参数、连续的中心对称参数以及积累中心对称参数随 y 轴坐标的分布图

中心对称参数在对称倾斜晶界中仅与晶界平面法线方向(即图 3.2 中的 y 轴方向)相关，因为原子的周期性排列，在 y 轴上具有相同相对距离的原子具有相同的周边环境。而相同的晶界在 y 轴上的相对分布也是确定的，仅有轻微的扰动导致的区别。因此通过中心对称参数在 y 轴上的分布可以在一定程度上代表整个晶界模型的原子空间分布，而使用该分布将能使三维的空间信息简化到二维(y 轴坐标和中心对称参数数值)信息，突出晶界附近的结构特征。表 3.2 给出了部分 α -Fe 对称倾斜晶界($\langle 110 \rangle$ 倾斜轴 $\Sigma 9$)的原子结构信息，可见中心对称参数在远离 y 坐标 0 值处有较小的中心对称参数，而在靠近 y 坐标 0 值处有较大的正数，并且具有相同的 y 坐标的几个不同 x 坐标、 z 坐标的原子具有相同的中心对称参数，因此通过中心对称参数在 y 轴上的分布来简化原子在三维空间中的分布是可行的。

表 3.2 α -Fe 的<110>倾斜轴 $\Sigma 9$ 对称倾斜晶界参数(部分)

原 子 id	x 坐标	y 坐标	z 坐标	CS 值	原 子 id	x 坐标	y 坐标	z 坐标	CS 值
92	1.4420	-9.9162	3.5044	0.0000	141	2.6524	0.6036	1.4854	0.1272
89	3.9148	-9.9162	3.5044	0.0000	142	0.1796	0.6036	1.4854	0.1272
85	3.9148	-9.9162	7.5424	0.0000	144	2.6524	0.6036	5.5234	0.1272
87	1.4420	-9.9162	7.5424	0.0000	145	0.1796	0.6036	5.5234	0.1272
131	0.6190	-8.7487	1.4854	0.0001	143	1.0397	1.7453	3.5044	0.0142
139	3.0918	-8.7487	1.4854	0.0001	153	3.5124	1.7453	3.5044	0.0142
137	0.6190	-8.7487	5.5234	0.0001	146	1.0397	1.7453	7.5424	0.0142
140	3.0918	-8.7487	5.5234	0.0001	158	3.5124	1.7453	7.5424	0.0142
133	2.2653	-7.5861	3.5044	0.0002	148	4.3229	2.9264	1.4854	0.0024
136	4.7381	-7.5861	3.5044	0.0002	149	1.8501	2.9264	1.4854	0.0024
127	2.2653	-7.5861	7.5424	0.0002	154	4.3229	2.9264	5.5234	0.0024
130	4.7381	-7.5861	7.5424	0.0002	155	1.8501	2.9264	5.5234	0.0024
129	3.9176	-6.4151	1.4854	0.0007	150	2.6812	4.0830	3.5044	0.0007
132	1.4449	-6.4151	1.4854	0.0007	151	0.2085	4.0830	3.5044	0.0007
135	3.9176	-6.4151	5.5234	0.0007	156	2.6812	4.0830	7.5424	0.0007
138	1.4449	-6.4151	5.5234	0.0007	157	0.2085	4.0830	7.5424	0.0007
122	0.6137	-5.2585	3.5044	0.0024	147	1.0290	5.2540	1.4854	0.0002
134	3.0865	-5.2585	3.5044	0.0024	162	3.5017	5.2540	1.4854	0.0002
116	0.6137	-5.2585	7.5424	0.0024	168	3.5017	5.2540	5.5234	0.0002
128	3.0865	-5.2585	7.5424	0.0024	152	1.0290	5.2540	5.5234	0.0002
118	2.2761	-4.0775	1.4854	0.0142	163	4.3281	6.4166	3.5044	0.0001
121	4.7488	-4.0775	1.4854	0.0142	164	1.8554	6.4166	3.5044	0.0001
126	4.7488	-4.0775	5.5234	0.0142	170	1.8554	6.4166	7.5424	0.0001
124	2.2761	-4.0775	5.5234	0.0142	169	4.3281	6.4166	7.5424	0.0001
123	1.4160	-2.9358	3.5044	0.1272	159	2.6784	7.5841	1.4854	0.0000
120	3.8888	-2.9358	3.5044	0.1272	160	0.2056	7.5841	1.4854	0.0000
115	3.8888	-2.9358	7.5424	0.1272	166	0.2056	7.5841	5.5234	0.0000
117	1.4160	-2.9358	7.5424	0.1272	165	2.6784	7.5841	5.5234	0.0000
113	0.7128	-1.7444	1.4854	2.6528	161	1.0306	8.7487	3.5044	0.0000
119	3.1856	-1.7444	1.4854	2.6528	171	3.5034	8.7487	3.5044	0.0000
125	3.1856	-1.7444	5.5234	2.6528	172	3.5034	8.7487	7.5424	0.0000
114	0.7128	-1.7444	5.5234	2.6528	167	1.0306	8.7487	7.5424	0.0000
175	1.9492	-0.5877	3.5044	9.6081	178	4.3272	9.9150	1.4854	0.0000
176	4.4220	-0.5877	3.5044	9.6081	179	1.8544	9.9150	1.4854	0.0000
173	1.9492	-0.5877	7.5424	9.6081	183	4.3272	9.9150	5.5234	0.0000
174	4.4220	-0.5877	7.5424	9.6081	184	1.8544	9.9150	5.5234	0.0000

由于各倾斜角晶界的晶界附近的中心对称参数的个数是不固定的，这不利于在接下来的模型中进行输入，因此通过局部加权二次回归模型^[44] (Locally Weighted Quadratic Regression, LWQR) 对离散的中心对称参数进行了拟合，得到连续的中心对称参数分布(如图 3.6 红色线所示)。相比于普通线性回归模型，局部加权回归模型的参数是变化的，它是依据预测点的位置来决定的，每一次预测时，按照一定的权重分布函数为所有训练样本赋予了全新的权重，然后据此更新了模型的参数后再进行最小均方差拟合得到预测值。局部加权回归通过这种方式提升了模型的拟合能力，同时还可以通过调整权重分布函数改变模型的拟合程度，得到最需要的预测结果。局部二次回归模型应用在中心对称参数和 y 轴坐标上的目标公式为：

$$\min \sum_{i=1}^m \omega^{(i)} \left[CS^{(i)} - \beta_0 - \beta_1 y^{(i)} - \beta_2 (y^{(i)})^2 \right]^2 \quad (3.5)$$

其中 m 为离散的中心对称参数的个数， $y^{(i)}$ 是对应的原子在垂直于晶界平面方向上的坐标， β_0 、 β_1 、 β_2 为模型参数，而 $\omega^{(i)}$ 是预先设置的权重系数，由预测点 y 和样本点 $y^{(i)}$ 之间的径向基核函数组成：

$$\omega^{(i)} = \exp\left(-\frac{(y^{(i)}-y)^2}{2k^2}\right) \quad (3.6)$$

对每一个预测点 y 进行预测时均会有一组相应的 $\omega^{(i)}$ 权重因子，调整衰减 k 可以得到最合适的拟合曲线， k 越大权值随距离下降得越快，在这里 k 设置为 3.3。通过最小二乘法可以对目标公式求解：

$$\boldsymbol{\beta}^T = (\mathbf{Y}^T \mathbf{W} \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{W} (\mathbf{CS}) \quad (3.7)$$

上式加粗符号代表矩阵， \mathbf{W} 为权重矩阵， \mathbf{Y} 为样本 y 轴坐标矩阵， \mathbf{CS} 为中心对称参数矩阵， $\boldsymbol{\beta}$ 为要求解的模型参数。确定好模型参数 $\boldsymbol{\beta}$ 后即可通过局部二次回归模型对待预测点求解，最终结果如图 3.6 中红色线。此时，离散的中心对称参数通过这种方式成为了连续的中心对称分布为之后在卷积神经网络中输入提供了便利。连续的中心对称参数对 y 轴求解二阶导数可以作为判断晶界区域和晶粒区域交界面的依据，如图 3.7 所示，二阶偏导数局部的峰值即为晶界区域和晶粒区域的交界面位置。对本文中所有的对称倾斜晶界所在区域进行了统计，如图 3.8 所示，最小的晶界区域在 y 轴上的区域范围在 $(-5\text{\AA}, 5\text{\AA})$ ，最大的晶界区域在 y 轴上的区域范围在 $(-20\text{\AA}, 20\text{\AA})$ 内，更多的晶界范围在 $(-11\text{\AA}, 11\text{\AA})$ 上。

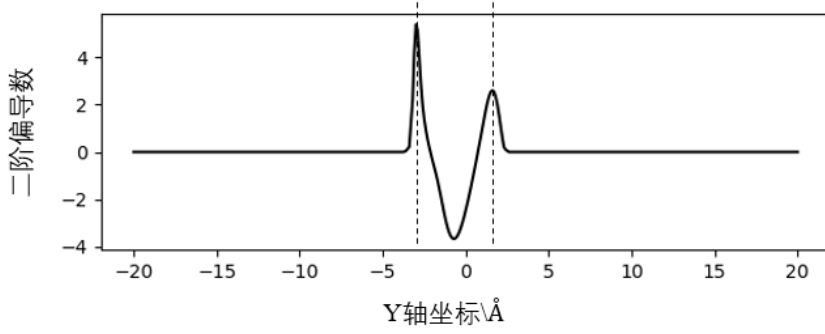


图 3.7 <110>倾斜轴的 $\Sigma 9$ 对称倾斜晶界上连续中心对称参数的二阶偏导数

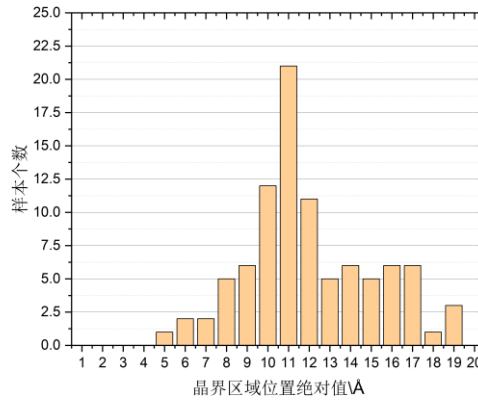


图 3.8 对称倾斜晶界晶界区间范围统计

除了通过局部权值二次回归构造出了连续的中心对称参数，本文还通过伊番科尼可夫核函数 (Epanechnikov Kernel) 融合进原子密度信息构造出了累加中心对称参数 (ACS, Accumulated Centro-symmetry Parameter, 如图 3.6 绿色线所示)。不同倾斜角的对称倾斜晶界在 y 方向上的原子密度是不同的，而连续的中心对称参数没有办法反映出这个特征，因此积累中心对称参数在每一个预测点处会以该点周边区域的离散中心对称参数出发，对每一个该区域的离散中心对称参数形成一个伊番科尼可夫核，最后全部叠加起来求得待预测点的值，以此凸显晶界区域原子在 y 轴上的分布情况，具体公式为：

$$ACS(y) = \sum_{D^{(i)}} CS^{(i)} [d^2 - (y - y^{(i)})^2] \quad (3.8)$$

其中 $D^{(i)}$ 是样本点 $y^{(i)}$ 位于区间 $(y-d, y+d)$ 的集合， d 为伊番科尼可夫核函数区间的宽度，这里设置为体心立方 Fe 的半个晶格长度。

图 3.8 中统计发现本文构建的所有对称倾斜晶界在 y 轴上全部集中在 $(-20\text{\AA}, 20\text{\AA})$ 区间内，因此选用该区间范围内的中心对称参数和积累中心对称参数作为主要的晶界结构特征。同时，为了便于在之后模型中作为输入，在该区间等间隔地对连续的中心对称参数和积累中心对称参数各取 401 个值作为晶界特征。

至此，对 α -Fe 对称倾斜晶界总共提取了三种晶界结构特征：径向分布函数、中心对称参数、积累中心对称参数。每种特征取 401 维。

3.3 设计 TextCNN 模型

3.3.1 TextCNN 模型框架

基于 2.2 章节，本文挑选了 TextCNN 模型^[53]作为基础模型。TextCNN 本身被应用于文本分类、机器翻译等任务中，通过不同类型的卷积核自动地对特征进行组合筛选从而获得丰富的语义特征。TextCNN 对文本等一维数据具有很好的特征提取能力，因此，本文将其应用于提取晶界的结构特征，设置了多个不同尺寸的卷积核并比较了各种尺寸卷积核提取的特征对模型预测能力的影响。如图 3.9 所示，TextCNN 模型的卷积核尺寸与寻常的卷积核不同，在一个维度上是固定的，这是由于它的输入数据一般是一维，并且卷积核只在一个方向上进行移动卷积，这很适用于本文在 3.2 章节中构造的晶界结构参数。

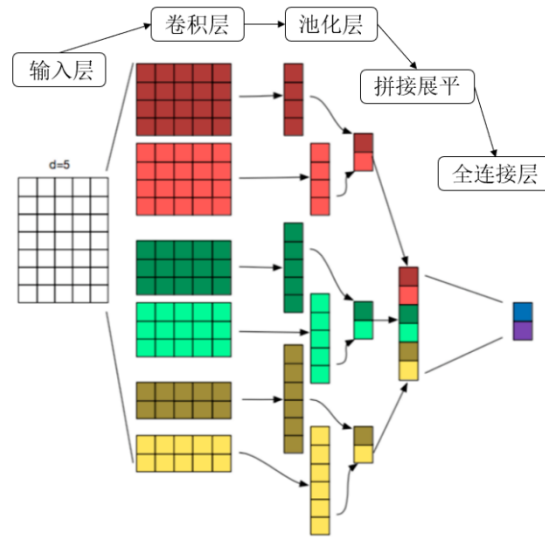


图 3.9 TextCNN 模型框架图^[53]

3.3.2 TextCNN 模型参数

经过模拟实验分析，本文通过 PYTORCH 1.5.1 搭建了相应的卷积神经网络模型，整个模型的参数量在 1.7×10^5 左右，基本的超参数如下表 3.3 所示，并在下一章节中，将通过预测集上的表现对参数造成的模型影响进行了比较分析。

表 3.3 TextCNN 超参数设置

超参数	值
输入向量大小	(401,1)

卷积核大小	(2,4,10,24,50,90)
每个尺寸卷积核个数	(50, 100, 150, 200)
激活函数	ReLU
池化策略	Chunk-Max-Pooling(划分为 4 个区间)
Dropout	0.5
全连接层隐层大小	(2400,64)
损失函数	smooth L1
批大小(batch)	10
迭代次数(epoch)	1000
优化器	Adam
学习率	3e-4, 1e-4, 1e-5

数据流动的整体路径如下图 3.10 所示，401 维的输入向量首先输入卷积层，并通过(2, 4, 10, 24, 50, 90)等特定尺寸的卷积核进行卷积计算，设置卷积步长为 1，分别得到(400, 398, 392, 378, 352, 312)维特征。随后，将卷积后的特征输入带有 ReLU 激活层进行非线性操作。再通过 Chunk-Max-Pooling 池化方法将各卷积层特征划分为四个区间，取出各区间中局部最大特征，按序拼接后输入拼接层。由于每种尺寸的卷积核设置为 $k(50, 100, 150, 200)$ 个，最终拼接后将得到 $6*k*4$ 维特征，将其展平(Flatting)后输入带有两层隐藏层(64, 16)的全连接层进行回归预测。在全连接层中，按一定概率暂时丢弃一定数量的神经单元^[31]，防止在卷积网络中可能出现的过拟合现象，丢弃概率称为“Dropout”，设置为 0.5。

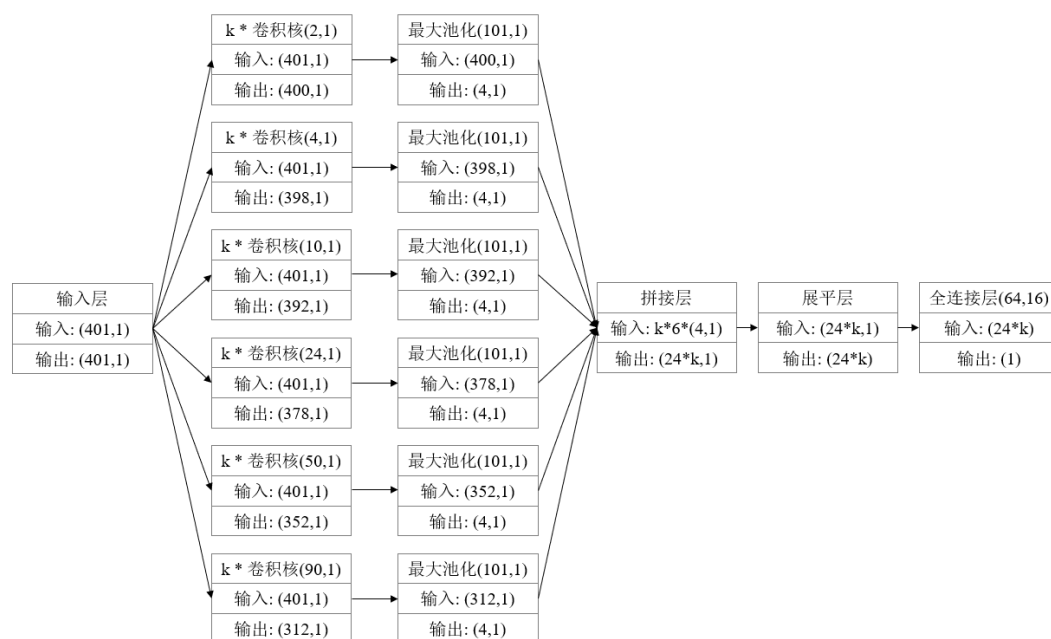


图 3.10 卷积神经网络模型数据流动示意图

表 3.3 还涉及了一些关于卷积模型结构和模型训练的超参数，于 2.2 章节中已经详细介绍过，这里不再展开。

3.3.3 实验环境

深度学习框架通过简单的接口可以实现基础的模型结构，研究者们还可以对其进行个性化设计，这为进行深度学习研究提供了便利。因此，随着时间的推移，目前有了多种深度学习框架在全世界范围内流行，其中，PyTorch 是以 Python 作为开发语言的深度学习框架，不仅能随时调整网络结构还能将整个网络搬运到 GPU 中进行加速计算，并能自动地进行求导，随时可以调试运行。PyTorch 相比于其他框架，代码更符合 Python 语言习惯，便于初学者进行学习。因此本文主要通过 PyTorch1.5.1 搭建了 TextCNN 模型。主要的环境配置如下表所示：

表 3.4 实验环境及配置

实验环境	环境配置
操作系统	Windows 10
CPU 型号	Intel core i7-9750H
GPU 型号	NVIDIA GeForce RTX 2060
编程语言	Python 3.6 + Pycharm
深度学习框架	PyTorch1.5.1

3.3.4 数据集预处理

(a) 按倾斜角分层抽样

为了能更加真实地了解模型在未知数据集上的表现，需要将所有样本库分为训练集和测试集，在训练集上进行模型的学习、验证，在测试集上进行模型的评估。从图 3.3 中可以看到相邻倾斜角的晶界能较为接近，因此本文通过按倾斜角分层抽样的方法划分训练集和测试集。首先统计了晶界在各倾斜角区间的样本个数，如图 3.11 所示， $\langle 100 \rangle$ 倾斜轴晶界分布相对均匀，而 $\langle 110 \rangle$ 倾斜轴晶界分布较为不均匀，在 $(15^\circ, 25^\circ)$ 、 $(65^\circ, 75^\circ)$ 、 $(105^\circ, 115^\circ)$ 、 $(135^\circ, 145^\circ)$ 、 $(155^\circ, 165^\circ)$ 等区间具有较小的个数。因此，本文按照各区间样本个数比例进行抽取测试集，尽可能地保留稀疏区间的样本，最终以 28 个 $\langle 100 \rangle$ 倾斜轴晶界和 50 个 $\langle 110 \rangle$ 倾斜轴晶界作为原始训练集，剩下的晶界作为测试集。

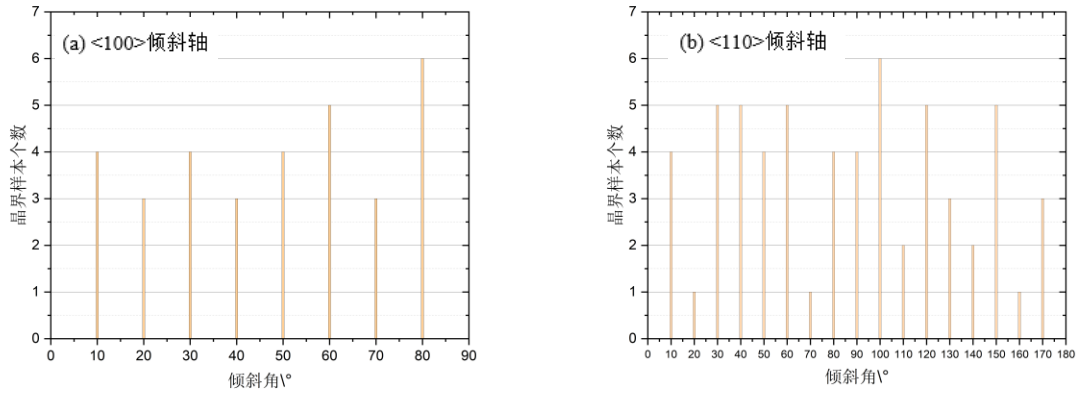


图 3.11 晶界在各倾斜角区间内样本个数统计(a)倾斜轴<100>(b)倾斜轴<110>

(b) 数据增强

由于通过 LAMMPS 构建晶界库需要花费大量的时间, 对称倾斜晶界库包含的晶界数量仍不够满足深度学习输入的需求, 较小的样本个数会导致模型的鲁棒性不足。因此在晶界相对分布不被改变的前提下, 本文通过数据增强^[57]的方法扩充了对称倾斜晶界样本库。数据增强是通过引入微小的改变以获得更多的数据, 常见的数据增强方式有翻转(Flip)、移位(Translation)、旋转(Rotation)、缩放(Scale)、裁剪(Crop)和噪声(Noise)等^[57]。能够进行数据增强的前提是样本增强前后具有不变性, 这就好比在光学字符识别任务中, 改变前后的字符需代表同一含义。需要注意的是只有训练集才能进行数据增强, 而测试集为了保证样本的真实性是不能够进行数据增强的。

参考图 3.6, 中心对称参数和积累中心对称参数是相对于 y 轴的分布, 主要表现的是一种相对分布的结构特征, 而绝对参考系的设置将不会改变整个晶界结构, 因此本文在 y 轴上对中心对称参数和积累中心对称参数进行了平移和翻转, 改变前后的特征代表都是同一晶界结构。最终将训练集上的中心对称参数和积累中心对称参数扩展为 1848 个<100>倾斜轴样本和 3300 个<110>倾斜轴样本。然而径向分布函数是相对于原子径向距离的分布, 翻转或者平移会改变晶界结构, 因此, 本文未对径向分布函数进行增强。在 LAMMPS 生成对称倾斜晶界的过程中, 同时会生成较多亚稳态的对称倾斜晶界, 这些晶界与具有最低晶界能的稳态晶界间的差异是原子位置具有较小的差异。不严格得说, 这相当于对稳定的晶界加入了噪声, 因此也可以通过这种方式进行数据增强, 但要保证的是, 引入的亚稳态晶界和稳态晶界间的差异不能过大, 否则晶界结构发生改变, 则破坏了数据增强前后样本不变的前提。下文中为了谨慎起见, 暂未引入亚稳态的晶界, 仅仅通过翻转和平移的方法增强了稳态对称倾斜晶界的中心对称参数和积累中心对称参

数特征。

3.4 预测晶界能结果

3.4.1 评价标准

在前一章节中提到为了准确地评估模型在未知晶界上的表现，需要将 α -Fe 铁的对称倾斜晶界库划分为训练集和测试集，在测试集上进行预测晶界能并与分子动力学计算的目标晶界能进行比较得到模型的预测效果。因此，在 TextCNN 模型的训练阶段，设置每 100 个训练周期输出一次测试集上的评价结果。具体的评价方法是计算测试集上所有对称倾斜晶界的模型预测晶界能和目标晶界能间的平均相对误差 $\bar{\varepsilon}$ ，具体公式为：

$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \frac{100|\gamma_i - o_i|}{\gamma_i} \quad (3.9)$$

式中 γ_i 为 Lammmps 计算的目标晶界能， o_i 为预测晶界能， n 为测试样本数。通过平均相对误差可以避免由对称倾斜晶界本身的晶界能差异造成的影响，这样每个样本都将具有同样的影响因子，可以更加准确地了解到模型在整个测试集上的预测效果。

3.4.2 预测晶界能结果

图 3.12 表示了中心对称参数和积累中心参数作为晶界特征在不同尺寸卷积核的 TextCNN 模型上预测晶界能的平均相对误差随训练周期的变化。这里同尺寸的卷积核个数均设置为 100 个。

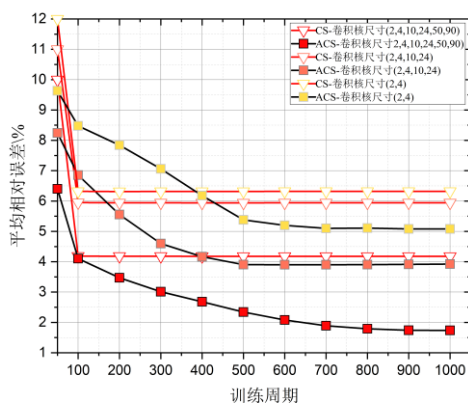


图 3.12 不同模型预测晶界能的平均相对误差随循环周期变化

可以看到，平均相对误差随循环周期逐渐下降，在 1000 个周期时均能趋于平稳，其中 ACS 特征在参数更新的学习率高于 CS 特征的情况下误差下降趋势

仍更缓慢，这在一定程度上反映了 ACS 特征相比于 CS 特征更加地复杂，难以学习。在所有模型中，ACS 特征在卷积核尺寸为(2,4,10,24,50,90)的网络中取得了最低的平均相对误差 1.74%，而同模型 CS 特征的平均相对误差仅为 4.17%，增大了 2.43%。在其余同卷积核尺寸模型中，ACS 特征模型的平均相对误差均低于 CS 特征模型。进一步分析同特征在不同尺寸卷积核模型上的表现：ACS 特征在 (2,4)、(2,4,10,24)、(2,4,10,24,50,90)卷积核模型上的平均相对误差依次为 5.08%、3.90%和 1.74%，CS 特征为 6.31%、5.94%和 4.17%。通过比较差值发现，无论是 ACS 特征还是 CS 特征，更大尺寸的卷积核使模型平均相对误差下降更明显。而由于输入参数规模的限制，进一步设置更大尺寸的卷积核较为困难，通过增加同尺寸卷积核的个数亦能得到更大规模的卷积网络模型，因此在图 3.13 中在卷积核尺寸为(2,4,10,24,50,90)的模型的基础上，通过改变单个尺寸卷积核的个数，比较了更大规模网络的平均相对误差。

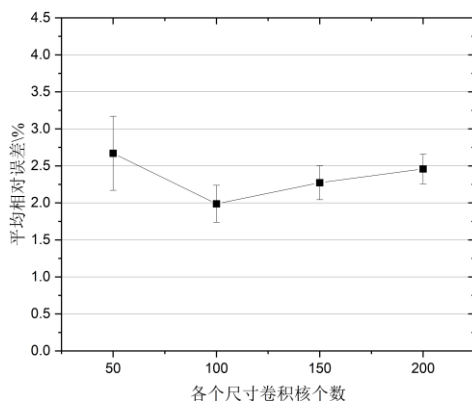


图 3.13 平均相对误差随卷积核个数的变化

结果可见，在单尺寸卷积核个数为 100 的模型上平均相对误差是最低的，稳定在 2%左右，最低可得 1.74%，而进一步增大个数使模型的平均相对误差增大，无法提升模型的预测效果。同时，根据图上各点误差棒的情况，可以看到在较小卷积核个数的时候模型预测结果的整体波动是较大的，而随着卷积核数目的增多，误差棒的宽度逐渐变小，模型也变得更加的稳定。但较多的卷积核个数反而减弱了模型预测的效果，具体上升的原因值得进一步的分析考证，在本文中尚未进行求证。

下面针对具有最低平均相对误差的(2,4,10,24,50,90)卷积核模型,进一步展示了分子动力学计算的晶界能和模型预测的晶界能间的差距。由于测试集样本数较少,在图中一并加入了由分子动力学直接生成的训练集晶界。首先是以中心对称参数作为特征的 TextCNN 模型的结果,如下图 3.14 所示:

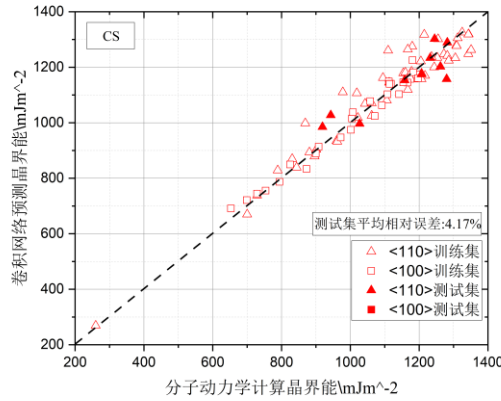


图 3.14 所有样本集上原始晶界能和预测晶界能对比图(CS)

图 3.15 是以积累中心对称参数作为特征训练模型的结果:

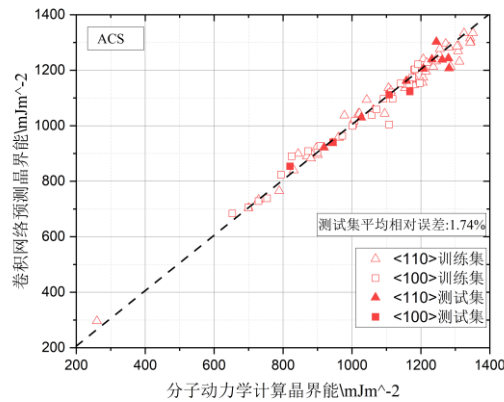


图 3.15 所有样本集上原始晶界能和预测晶界能对比图(ACS)

综合图 3.14 和图 3.15 可见,由于 CS 特征训练的 TextCNN 模型在测试集上的平均相对误差(4.17%)高于 ACS 特征训练的模型(1.74%),CS 特征的模型各样本晶界能的分布更为离散,出现较大相对误差的可能性也更大,肉眼可见地 ACS 特征的模型的样本点更集中在预测晶界能等于计算晶界能的虚线附近。为了能更好的对样本点离散程度进行判断,图 3.16 统计了各预测结果的分布情况:

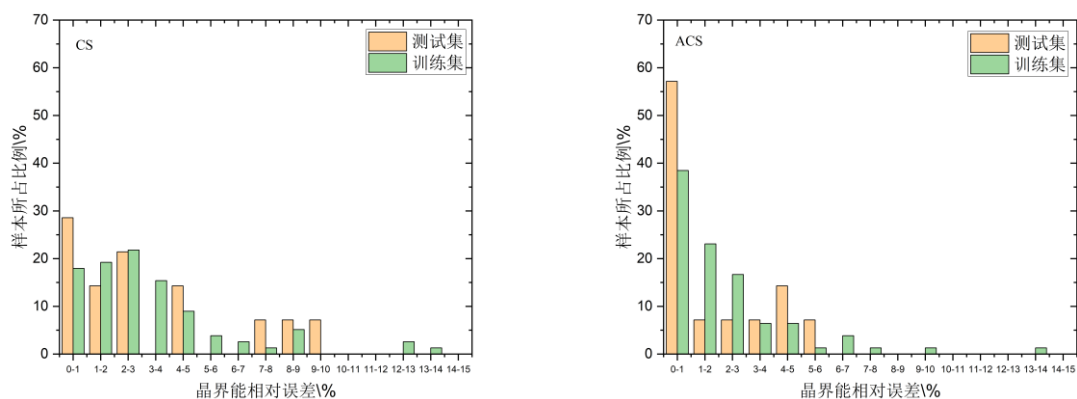


图 3.16 预测结果统计分布图

从上两图中可见,ACS 特征训练的模型结果相对于 CS 特征的模型更为集中。并且在测试集上,ACS 特征的模型最大相对误差在的 5-6%区间内,而 CS 特征的模型最大相对误差则达到了 9-10%,同时 ACS 在 0-1%区间内的样本所占比例达到了 57%,而 CS 在该区间的样本所占比例仅为 28%,可见,ACS 特征训练的模型有更为优秀的预测效果,在训练集上也有与测试集相似的结果。

综合以上的表现,ACS 作为晶界特征在相同的 TextCNN 模型中有比 CS 特征更优的效果,在卷积核尺寸为(2,4,10,24,50,90)、同尺寸卷积核个数为 100 的 TextCNN 模型中预测晶界能有最低的平均相对误差 1.74%。

图 3.17 表现了以径向分布函数作为晶界特征通过卷积核尺寸为 (2,4,10,24,50,90),同尺寸卷积核个数为 100 的 TextCNN 模型在预测集上预测晶界能的表现随训练周期 epoch 的变化。之所以未将径向分布函数与中心对称参数和积累中心对称参数一起作为晶界特征进行比较是因为对 ACS 和 CS 特征通过数据增强方式将其训练集进行了扩充,而径向分布函数因为本身含义的限制,在

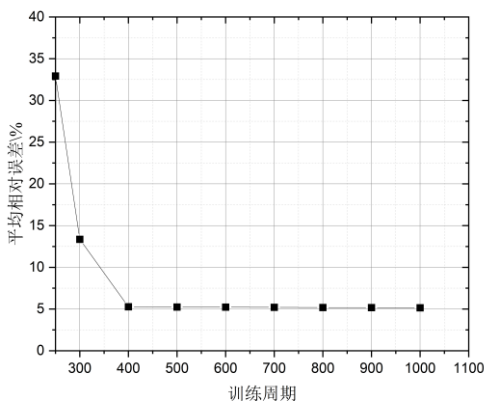


图 3.17 径向分布函数作为晶界特征的模型平均相对误差随循环周期变化

径向距离上不能够进行翻转和移动。因此，径向分布函数作为晶界特征的训练集尺寸仅有 50 个<110>倾斜轴对称倾斜晶界和 28 个<100>倾斜轴对称倾斜晶界。

可见，径向分布函数作为晶界特征在该 TextCNN 模型上最好的表现仅为 5.15%，远不及 ACS 作为晶界特征在相同的模型上的表现。这很可能是因为训练集个数太少导致的，由于训练集个数的限制，径向分布函数难以体现其表现晶界结构的能力。相比于 Joshua A^[24]利用晶界区域的径向分布函数通过机器学习方法在 Al 的非对称倾斜晶界上预测晶界能取得的结果(晶界能在 200 至 600mJ/m² 范围内的平均绝对值误差为 11.35 mJ/m²)，利用径向分布函数结合 TextCNN 模型在 Fe 的对称倾斜晶界上的预测结果更差。而 ACS 作为特征结合 TextCNN 模型在预测晶界能任务上却取得了更优的结果。

3.4.3 分层抽样和数据增强对模型的影响

为了进一步确认按倾斜角分层抽样对模型预测能力的影响是正向的，在部分的数据集上分别以中心对称参数和积累中心对称参数作为晶界特征在卷积核尺寸为(2,4,10,24,50,90)、同尺寸卷积核个数为 100 的 TextCNN 模型上，比较了三种抽样方式(按倾斜轴分层抽样、随机抽样、不均匀抽样)在相同的测试集上取得的预测效果。其中，不均匀抽样是只选择了倾斜角较大的样本，而随机抽样是在所有晶界中随机得进行抽样。结果如图 3.18 所示，不论是 ACS 特征还是 CS 特征，分层抽样得到的测试集平均相对误差均是最小的，随机抽样次之，不均匀抽样最大。这表明了不同倾斜角的对称倾斜晶界的晶界结构差异是较大的，仅适用一定倾斜角范围内的晶界可能导致样本在特征空间的分布不均匀，模型易发生过拟合的情况，在更大倾斜角范围内的晶界上预测结果较差。因而通过按倾斜角分层抽样的途径，能使训练样本空间分布得更加均匀，使模型在测试集晶界上能有更准确的预测结果。

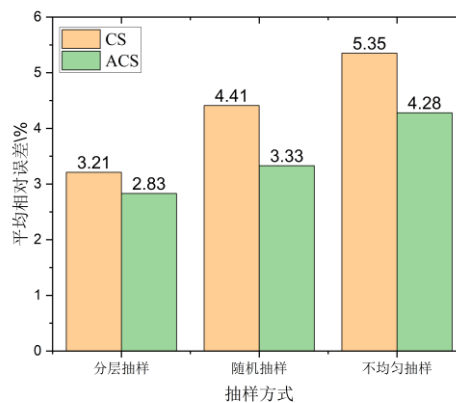


图 3.18 不同抽样方法在 TextCNN 模型上预测晶界能的平均相对误差的变化

在 3.3.4 章节中, 通过数据增强的方式对中心对称参数和积累中心对称参数进行了一定的扩充。在本节中, 将进一步研究数据增强的样本个数对模型预测效果的影响。在初识卷积神经网络时, 由于卷积操作的特性, 卷积神经网络被认为拥有平移不变性^[27], 但在实际任务中, 研究者通过翻转、平移、缩放、扭曲等强相关性的几何变换和像素扰动、添加噪声等方式进行的数据增强的的确确提升了卷积神经网络的效果^[57], 这说明了卷积神经网络的平移不变性是不完全的。数据增强更可能是从数据角度出发对模型进行了正则化操作, 这与 Dropout 等形式引入的正则化一样, 能够减小模型的结构风险, 提高模型的鲁棒性。因此本文期望通过平移翻转等强相关性操作来提高模型效果的想法是合理的。但在引入数据增强样本的过程中, 很可能顺带引入了一些经验风险, 当风险过大时, 样本空间分布出现了改变, 则会对模型的预测能力产生负作用, 这就要求数据增强前后的样本的主要特征还是保持一致。本文通过平移和翻转的方式对 CS 和 ACS 特征进行了数据增强, 即使这种方式只是简单的几何变换, 但随着平移尺度的增加, 可能也会给模型带来一定的经验风险。因此, 在图 3.19 中, 以积累中心对称参数作为晶界特征在卷积核尺寸为(2,4,10,24,50,90)、同尺寸卷积核个数为 100 的 TextCNN 模型上, 通过改变增强样本的个数(同时改变了平移尺度)观察模型在预测集上的平均相对误差的变化。

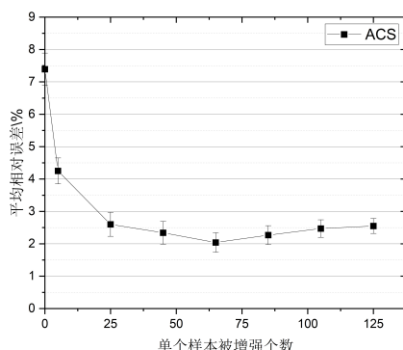


图 3.19 不同数据增强样本个数下预测晶界能平均相对误差的变化

上图结果表明, 相较于未进行数据增强的模型, 合理地引入增强样本的确能够降低模型预测晶界能的平均相对误差, 随着增强样本个数的增多从未进行数据增强的 7.4%到增强样本个数为 65 左右的最低点 2.0%, 在该最低点后, 模型的平均相对误差出现了一定的上升。由此可见, 平移和翻转的数据增强方式在一定范围内为该模型提高了预测效果, 但随着平移尺度的加大和增强样本个数的增多, 对引入了一些负面的噪声, 改变了训练数据的主要特征分布, 导致模型的预测效果产生了一定的下降。因此, 选择在合适的范围内的增强尺度和增强样本个数对提升模型预测性能是十分重要的。

3.4.4 与传统机器学习方法预测结果对比

为了更好地展现卷积神经网络在预测 RAFM 钢晶界能任务上的优势，本文对比了两种机器学习方法在现有的晶界特征基础上预测晶界能的效果。机器学习模型预测晶界能的流程与卷积神经网络预测晶界能的流程是一致的，只需要将卷积神经网络替换成相应的机器学习模型即可。两种机器学习模型分别是支持向量机(SVM)^[58]模型(2.1 章节中已介绍)和 1.4 章节中介绍的多层感知机模型(MLP)。下面先对这两模型的参数设置进行一个说明。

(a)支持向量回归模型

利用 Scikit-learn 0.23.2 模块在 Python 3.6 平台上实现了支持向量回归模型，并分别采用了中心对称参数和积累中心对称参数作为晶界特征。该模型拥有一些参数，其中惩罚系数“C”是对公式 2.7 中的 loss 增加了系数，一般取值为 10 的 n 次幂，“C”越大得到的 loss 越大，模型将更符合训练集的分布，但过度地符合训练集分布将导致模型在预测空间的性能有所下降；惩罚距离“epsilon”与原理中提到的预先设置的偏差 ϵ 对应；通过“kernel”可以指定核函数类型；目标误差“tol”代表损失函数达到一定的数值时模型结束训练。详细的模型参数如下表 3.5 所示：

表 3.5 支持向量回归模型参数

晶界特征	中心对称参数	积累中心对称参数
惩罚系数 C	1.0	0.1
惩罚距离 epsilon	1.0	0.1
核函数 kernel	径向基核函数	1 阶多项式核函数
目标误差 tol	0.001	0.001

(b)多层感知机模型

同样是使用 Scikit-learn 0.23.2 模块在 Python 3.6 平台上实现了具有两层隐藏层的多层感知机模型。以均方误差作为损失函数，优化器选用 Adam，激活函数使用修正线性单元(ReLU)，最大迭代步数设置为 50000 步。对于 CS 晶界特征，隐藏层单元个数设置为(64,16)，对于 ACS 晶界特征隐藏层单元个数设置为(128,16)。

由于 SVM 算法并非迭代算法，因此不能像 TextCNN 模型一样观察测试集上预测晶界能的平均相对误差随循环周期数 epoch 的变化。这里通过逐渐增大训练样本数的方式观察两模型预测能力随训练样本数的变化。使用倾斜轴为<110>的 50 个对称倾斜晶界作为训练样本，依次从中随机抽取 5 个样本加入训练集，同时每个样本通过平移和翻转的数据增强方式扩展到 66 个。最终在相同的测试集

上的结果如图 3.20 所示：

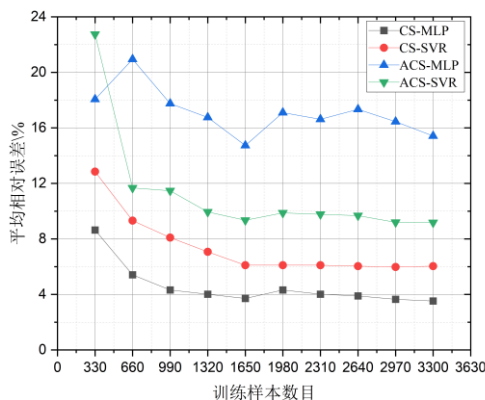


图 3.20 SVR 和 MLP 模型预测能力随训练样本数的变化

可见，各模型的平均相对误差随训练样本数的增加总体上呈现的是逐渐下降的过程，在训练样本数为 990 前下降的趋势较明显，随后变得较缓慢。值得注意的是在训练样本数为 1650 时，在多个模型上出现了较低的平均相对误差。出现该现象的原因可能是随机抽取训练集时在 1650 样本个数的训练集已经取到了倾斜角范围分布较全面的样本空间了。此时训练集样本空间和测试集样本空间的分布更为一致，而随着进一步增加训练样本个数，反而使模型出现了过拟合的现象，使得模型的预测有一定的下降。导致该问题的最根本原因还是通过分子动力学构建的对称倾斜晶界数仍旧太少，测试集数目少，测试样本空间分布难以完全代表未知晶界空间分布，因此，想要更精确的判断各模型的性能，仍需要进一步扩充对称倾斜晶界样本库，在这里仅以有限的样本个数对各模型的预测效果进行初步的判断。

从两晶界特征来说，可见中心对称参数在两机器学习模型上的预测效果均比积累中心对称参数来得更好，这与 TextCNN 模型的预测结果相反。究其原因，积累中心对称参数融合了原子密度信息，在提取信息能力较弱的机器学习模型中得不到很好的表现，而 TextCNN 模型凭借其优秀的提取空间信息的能力，很好得提取了积累中心对称参数中隐含的结构信息，为预测晶界能的任务提供了强大的支撑。

综合来看，中心对称参数在 MLP 模型上预测结果稳定在 3.5%左右、在 SVM 模型上的预测结果为 6.0%左右，与 TextCNN 模型上的预测效果 4.71%相比，MLP 模型是最合适中心对称参数，较 TextCNN 模型提升约 1.2%左右。而积累中心对称参数在 MLP 模型上的预测结果为 15%左右、在 SVM 模型上的预测结果为 9.2%

左右，与在 TextCNN 模型上的预测效果 1.74%相比，TextCNN 模型对积累中心对称参数预测晶界能效果的提升是十分显著的，TextCNN 模型比 MLP 模型提升了约 13.2%左右。这与计算机算法中常说的没有免费的午餐定理(No Free Lunch Theorem)^[59]一致，没有一劳永逸的算法，具体问题需要具体分析，更简单的算法也有可能特定的问题上有更为优秀的表现。

3.5 本章小结

本章节首先介绍了利用 TextCNN 模型预测对称倾斜晶界的晶界能的总体流程，构建了 α -Fe 对称倾斜晶界库，提取了相应的晶界特征(径向分布函数、中心对称参数和积累中心对称参数)，构建了相应的 TextCNN 模型。最后对 TextCNN 模型预测晶界能的效果进行了全面的分析。通过比较实验发现，相比于 CS 特征和 RDF 特征，ACS 作为晶界特征在卷积核尺寸为(2,4,10,24,50,90)、同尺寸卷积核个数为 100 的 TextCNN 模型中有最佳的预测效果，平均相对误差为 1.74%，相比于机器学习方法有较大的提升。同时，通过对比实验，验证了按倾斜角分层抽样和数据增强的方法对模型的预测效果具有提升效果。至此，通过第 3 章节对整个 TextCNN 模型预测 α -Fe 对称倾斜晶界的晶界能任务有了详尽的介绍。

第 4 章 基于长短时记忆网络研究间隙氦在晶界中的迁移

1965 年, Barnes^[60]首先在 Nature 上发表了关于不锈钢和镍基合金在高温下受到中子辐照后由于氦原子聚集而导致的脆化效应。实验表明, 当材料中的氦原子浓度达到了一定的程度后, 会在晶界处形成纳米尺度的氦泡, 从而材料具备了高温脆性。氦泡对材料的影响在聚变堆中更为突出, 因为聚变堆的中子通量很高, 导致结构材料发生嬗变反应的几率很大^[61], 材料中的辐照氦效应显著。因此, 研究者们通过多种方法研究氦泡的成核生长机制, 包括基于量子力学的从头算、分子动力学和蒙特卡洛等多尺度模拟方法^[9]。本文聚焦于数据驱动的方法, 在分子动力学算例的基础上应用深度学习研究方法研究氦原子迁移的微观机制。主要的内容是利用循环神经网络算法对分子动力学构建的间隙氦原子在 α -Fe 中迁移的样本进行学习, 判断氦原子是否发生迁移, 并通过调整输入特征来研究影响氦原子迁移的主要因素。该流程与预测 α -Fe 对称倾斜晶界的晶界能任务相类似, 因此下文同样依据建立样本库、提取特征参数和建立模型的顺序进行介绍, 最终对该模型的判别结果进行了分析。整个模型的流程图如下图所示:

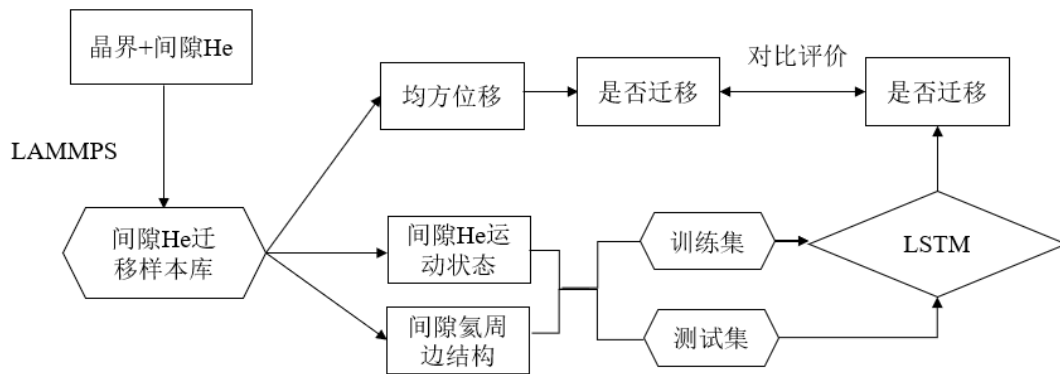


图 4.1 判断间隙氦是否迁移流程图

4.1 建立间隙氦迁移样本库

通过分子动力学模拟氦原子的迁移过程需要以下几个步骤：1.首先通过 2.2 节中在 Lammmps 中建立晶界的方法建立起 α -Fe 的对称倾斜晶界模型(倾斜轴 $\langle 110 \rangle$ 的 $\Sigma 3\{112\}$ 晶界、 $\Sigma 11\{113\}$ 晶界和 $\Sigma 11\{332\}$ 晶界), 同时建立了完整的 α -Fe 晶粒模型作为对比, 建立完美晶粒的方法与建立对称倾斜晶界的方法类似, 这里不做过多的介绍; 2.在晶界模型中心加入间隙氦原子, 并为模型设定初始温度 T(完美晶粒和 $\Sigma 3$ 晶界为 350K, $\Sigma 11$ 晶界为 500K)。通过设定温度, 给予所有粒子初始速度, 配合粒子的初始坐标, 通过牛顿运动方程确定体系内包含氦原子在内的所有原子的运动状态。3.在等温等压(NPT)系综下进行演化, 计算每一个原子的新

的速度和新的坐标。若此时所有原子构成的系统温度与预先设置的温度出现偏差,则需要对原子的速度进行矫正。通过原子速度计算系统温度的具体公式如下:

$$T_{\text{system}} = \frac{\sum_{i=1}^N m_i v_i^2}{3Nk_b} \quad (4.1)$$

其中 N 为模型中总原子个数, T_{system} 为系统温度, k_b 为玻尔兹曼常数, m_i 为各个原子的质量, v_i 为该原子的运动速度由 $\sqrt{v_{x,i}^2 + v_{y,i}^2 + v_{z,i}^2}$ 计算得到。通过对原子速度乘上矫正因子使得系统的总温度符合设定, 矫正因子 f 的计算公式如下:

$$f = \sqrt{\frac{T}{T_{\text{system}}}} \quad (4.2)$$

最终将得到该温度下的氢原子在当前晶界结构中的迁移状态, 按一定的时间间隔(10fs)输出了一系列时间步(时间尺度 1ns, 样本个数 100000 个)下模型中氢原子的运动状态和所处环境。随后, 通过分析氢原子的位置变化来判断其是否发生了迁移, 为这些样本打上标签。实际操作中往往通过两种方法判断间隙氢是否迁移, 第一种方法是直接观察 LAMMPS 输出的 dump 文件, 很直观得对间隙氢是否迁移进行判断, 第二种方法是计算该时刻氢原子距离初始位置的绝对距离的平方的变化(均方位移, Mean Square Displacement, MSD)来判断是否发生迁移。值得注意的是, 这种判断方法是不适用于以氢原子初始距离为球心的球面上的迁移的。因此, 可以再设置一个初始位置, 计算各时刻氢原子距离该位置的绝对距离的平方的变化, 与之前的 MSD 共同作为判断原子是否发生迁移的依据。只要这两个距离变化中有一个出现跃迁的现象就判断此时的原子已发生迁移。如图 4.2 所示:

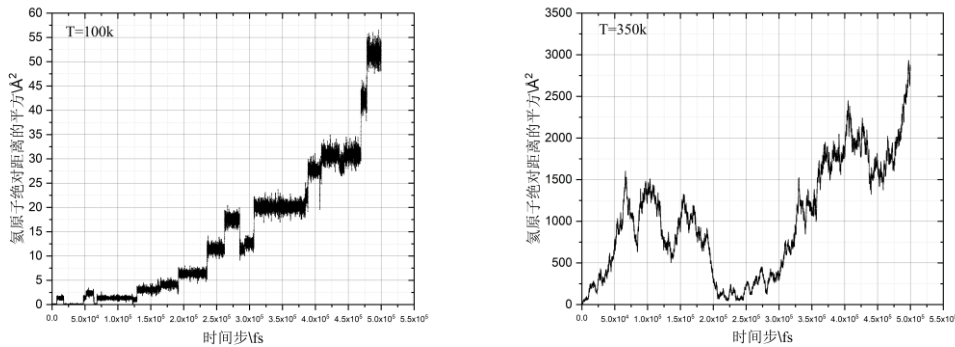


图 4.2 不同温度下间隙氢原子在 α -Fe 晶粒上的 MSD 变化

通过上图可以清晰的看出, 氢原子发生迁移时该距离的平方会出现一个跃迁现象, 在这个跃迁过程中的所有点均可以被认为是已经发生了迁移的点。上左图是在 100K 的温度下氢原子迁移的情况, 随着温度的增加, 氢原子热运动更为剧

烈，更容易发生迁移，跃迁的点增多，跳跃的幅度增大，平台期变得更加模糊，这为判断氢原子是否迁移造成了难度，如上右图温度为 350K 的情况。为了解决该问题，本文通过计算一个时刻前后 n 个时刻的氢原子相对于初始位置的距离和的绝对差值来判断该时刻氢原子是否迁移，得到的结果如图 4.3 所示：

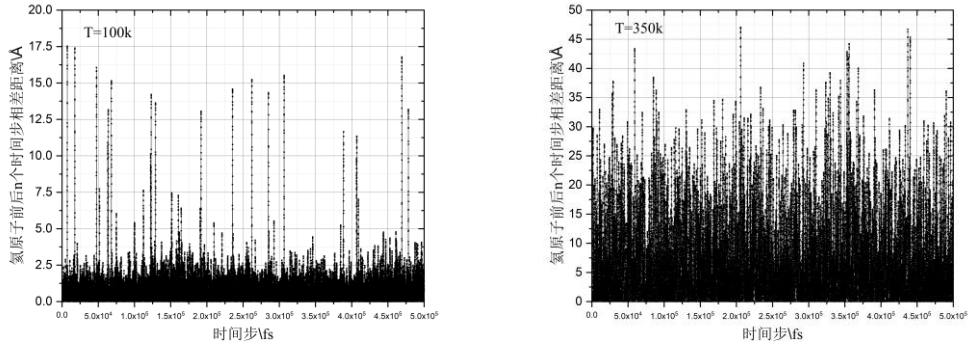


图 4.3 不同温度下氢原子距离和的绝对差值随时间步的变化

从上左图可见，100K 的温度下，在迁移时刻的前后 n 个时间步的距离和的绝对差值是较大的，而在未发生迁移的时间步上，该差值相对较小。通过这种方法，能缓解各平台期波动幅度差别较大带来的判别问题。但随着温度的增加，该方法也会由于氢原子剧烈的热运动而效果下降。如上右图中温度为 350K 时的变化情况，在平台期上也有较大的差值，不过相对于均方位移有所好转。因此在挑选样本输入循环神经网络时，应尽可能挑选前后 n 个时间步的距离和的绝对差值更为极端的样本(迁移样本挑选该差值较大的时刻，未迁移样本挑选该差值较小的时刻)。

总结：通过分子动力学方法建立了间隙氢原子在 α -Fe 的完整晶粒、倾斜轴 $\langle 110 \rangle$ 的 $\Sigma 3 \{112\}$ 对称倾斜晶界、 $\Sigma 11 \{113\}$ 对称倾斜晶界和 $\Sigma 11 \{332\}$ 对称倾斜晶界中迁移 1ns 的模型，其中完美晶粒和 $\Sigma 3$ 晶界的温度为 350K， $\Sigma 11$ 晶界的温度为 500K，以 10fs 的时间间隔对每个模型输出 100000 帧氢原子迁移状态。随后通过三种判断迁移的方法对样本进行标注。

4.2 提取氢原子迁移特征参数

为了更加准确地设计合适的氢原子迁移特征参数，首先需要对氢原子在金属中的迁移情况进行一个初步的认识。氢原子在金属中有五种迁移方式，包括间隙机制、空位机制、解离机制、替代机制和自间隙机制^[62]。其中，间隙机制是氢原子通过特定的间隙位置进行迁移的方式，间隙氢的迁移能很低，在 α -Fe 中仅有

0.08eV^[62]，很容易发生迁移，因此本文是基于间隙氦来进行研究的。

通过 Gao F^[63]在 2006 年发表的文献可知，间隙氦原子的迁移主要受到两个因素的影响：体系温度和晶界结构。体系温度的升高会加剧氦原子的热运动，更多无规则的热运动导致间隙氦原子更容易发生迁移。同时，由于间隙氦原子容易与各种缺陷进行结合，紊乱的晶界结构也对氦原子的迁移形成了重要的影响。下面分别从这两个因素出发设计氦原子迁移特征。

温度直接影响了氦原子的运动状态，因此代表温度效应的迁移特征可以直接从氦原子的运动状态入手。首先是氦原子的速度大小，本文考虑了氦原子在晶界模型参考系上 x、y、z 三个方向的速度分量(v_x , v_y , v_z)和总速度大小(v_{He})，并将其分别作为运动速度特征输入循环神经网络。同时，Gao F^[64]在 2009 年发表的关于 α -Fe 中空位、间隙氦、氦空位团簇迁移机制的文献中提到，间隙氦在不同的晶界结构中有各自特定的迁移方向，可见速度的方向也是影响氦原子迁移的重要参数之一。因此，本文计算了氦原子速度方向与晶界模型的 x、y、z 三个方向间的夹角(θ_x , θ_y , θ_z)，并以此作为氦原子的运动方向参数输入网络，其计算公式为：

$$\cos(\theta_i) = \frac{v_i}{\sqrt{v_x^2 + v_y^2 + v_z^2}} \quad (4.3)$$

晶界附近的氦原子迁移与晶界的结构有着重要的关系，想要提取晶界材料的结构参数可以参考预测晶界能任务中用到的中心对称参数，不同的是，这里是以氦原子作为中心原子，因此我们将其称为铁氦结构参数。从公式 2.4 中可以看到，计算中心对称参数需要提前设置近邻原子个数，对于 BCC 晶体来说一般设置为 8。本文在该基础上，同时计算了近邻原子个数为 2、4、6、8、10、12 的氦原子近邻铁原子的矢量和的平方，并将其一同作为氦原子的局部环境特征输入模型。

总结：氦原子迁移的特征参数总共挑选了 13 个参数，其中包括 3 个氦原子在三个方向上的速度分量(v_x , v_y , v_z)、1 个氦原子总速度大小(v_{He})、3 个氦原子的速度方向夹角(θ_x , θ_y , θ_z)、6 个氦原子的铁氦结构参数(CS_2 , CS_4 , CS_6 , CS_8 , CS_{10} , CS_{12})。在 4.4.3 章节中还将对比分析这些参数分别对模型的影响。

4.3 设计 LSTM 模型

依据 4.2 章节中设计的氦原子迁移的特征参数，在基础的 LSTM 模型上设计了适合该任务的分类模型。对速度分量(v_x , v_y , v_z)、速度夹角分量(θ_x , θ_y , θ_z)和铁氦结构参数(CS_2 , CS_4 , CS_6 , CS_8 , CS_{10} , CS_{12})这三类输入特征，分别将其输入各自的全连接层提取信息。再将提取出来的信息与速度总量(v_{He})一起同时输入 LSTM 模型中进行学习，得到 LSTM 的输出特征 h_t (这里设置成单层 24 维)，最终通过全连接层将 LSTM 的输出空间映射到两类迁移类别(迁移和未迁移)，随后利用 softmax 函数^[66]进行了归一化，得到了标准的迁移概率，并据此判断氦原子

是否发生了迁移。具体的模型结构如下图 4.4 所示：

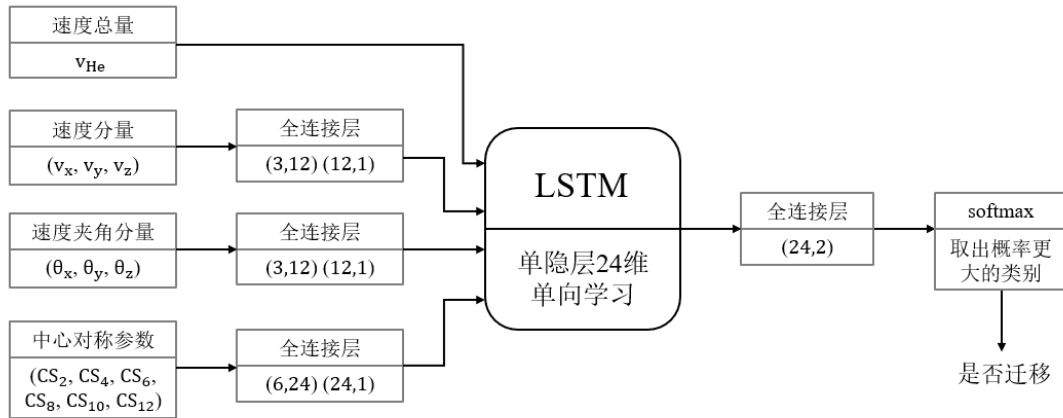


图 4.4 判断氦原子是否迁移 LSTM 模型示意图

其中 LSTM 模型的训练参数如下表 4.1 所示：

表 4.1 LSTM 模型训练参数

超参数	值
时间间隔	41
Dropout	0
损失函数	交叉熵
批大小(batch)	10
迭代次数(epoch)	50
优化器	Adam
学习率	3e-4

可以看到，由于 LSTM 需要输入的是一段时间间隔的特征参数，因此这里选择了当前时刻之前连续的 40 个时间间隔样本与它本身组成的 41 维样本序列作为一个完整的氦迁移样本输入特征，其中，每个样本的时间间隔为 10fs。由于判断氦原子是否迁移是一个分类问题，因此该模型以交叉熵(Cross Entropy)作为损失函数。交叉熵在二类别分类问题上具体的计算公式如下：

$$\text{loss} = -\sum_{i=1}^m [P(x_i^0)\log Q(x_i^0) + P(x_i^1)\log Q(x_i^1)] \quad (4.4)$$

其中 m 为样本个数， P 为模型预测氦原子迁移的概率分布， Q 为真实氦原子迁移的概率分布， x_i^0 表示该样本被判断为未发生迁移，而 x_i^1 表示发生迁移。迭代次数较小的原因是该模型的参数量较小，容易较快地发生收敛，其余的参数在 2.2 章

节中均已介绍过，这里不做重复地讨论。

4.4 各晶界中模型结果分析

4.4.1 评价标准

对于分类问题仅仅统计预测结果的准确性是不够的，而且我们重点关注的是发生迁移的样本。因此，可以通过查准率(precision)来统计预测为发生迁移的样本真实迁移的概率，还可以通过查全率(recall)来统计真实发生迁移的样本被预测为发生迁移的概率，结合这两种评判标准可以综合地分析判别模型的有效性。查准率 precision 的计算公式为

$$\text{precision} = \frac{TP}{TP+FP} \quad (4.5)$$

其中，TP(True Positive)表示真实迁移样本预测为发生迁移的个数，FP(False Positive)表示真实未发生迁移样本预测为发生迁移的个数。查全率 recall 则定义为：

$$\text{recall} = \frac{TP}{TP+FN} \quad (4.6)$$

其中，FN(False Negative)表示真实发生迁移的样本预测为未发生迁移的个数。可以看到查准率和查全率之间存在一定的关联，比如设置较低判断标准，样本很容易被判断为正例，但也会增大负例样本混入的可能性，这时就会有较高的查全率和较低的查准率。而如果设置比较高的判断标准，正例样本则很容易被抛弃，那就会有较高的查准率和较低的查全率。因此常用 F1 分数作为两者的折中点被用作评判标准，其计算公式如下：

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4.7)$$

本文在测试集上综合分析了该氦原子迁移分类模型关于发生迁移样本的查准率、查全率和 F1 分数。

4.4.2 氦原子迁移分类模型在各晶界上的结果分析

(a) α -Fe 完整晶粒，350K，演化时间 1ns

利用 2.2 章节中的氦原子迁移模型，在 1ns 的演化时间内，每 10fs 输出了一帧 2.1 章节中提到的氦原子迁移特征，一共得到了 10 万个氦原子迁移样本，通过计算一个时刻前后 20 个时刻的氦原子相对于初始位置的距离和的绝对差值，得到了如图 4.2 中 350K 的结果。为了保证对氦原子迁移样本分类的准确性，本文通过三种判断间隙氦是否迁移的方法综合挑选了迁移样本和未迁移样本各 10000 个样本。每个样本以自身时刻之前的连续的 40 个样本的氦原子迁移特征加上本身时刻的氦原子迁移特征共同组成了该时刻 LSTM 网络的输入特征。打

乱后随机选择 16000 个样本作为训练样本，4000 个样本作为测试样本。通过 LSTM 对训练样本进行了训练，训练集上的平均交叉熵随训练周期的变化如下图 4.5 所示：

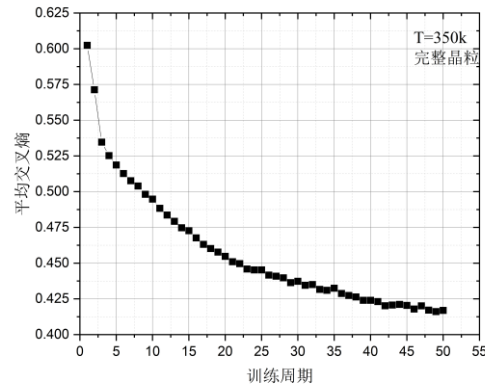


图 4.5 完整晶体模型训练平均交叉熵随训练周期的变化

可以看到，模型在训练完 50 个 epoch 后已经基本趋于收敛，因此在测试集上输出了该模型的评价指标，如下表 4.2 所示：

表 4.2 LSTM 模型在训练集上评价指标

输入特征	查准率	查全率	F1 分数
四类特征	0.82	0.83	0.82
无(v_{He})	0.84	0.82	0.83
无(v_x, v_y, v_z)	0.81	0.84	0.82
无($\theta_x, \theta_y, \theta_z$)	0.82	0.83	0.82
无铁氢结构参数	0.82	0.79	0.80
仅铁氢结构参数	0.69	0.72	0.70

具体在测试集中有 2068 个迁移样本，其中被判断为迁移样本的有 1716 个，剩下的 352 个未被正确判断，未迁移样本有 1932 个其中 1555 个样本被正确判断，而 377 个样本判断为迁移样本。通过该结果可以得到，以 LSTM 作为基础模型，通过这四种氢迁移参数是能够具备判断氢原子在完整晶粒中是否发生迁移的能力的，而且判断准确度达到了 80% 的水准。

在上表中还同时评价了模型在分别剔除 1 个氢原子总速度大小(v_{He})、3 个氢原子在三个方向上的速度分量(v_x, v_y, v_z)、3 个氢原子的速度方向夹角($\theta_x, \theta_y, \theta_z$)和 6 个氢原子的铁氢结构参数($CS_2, CS_4, CS_6, CS_8, CS_{10}, CS_{12}$)后进行模型训练并在测试集上预测的结果。本文尝试通过这种方式来分析各类氢迁移特征对模

型分类效果的影响。下面依次对四种参数进行对比分析：

去除氦原子总速度大小(v_{He})得到了更高的查准率，这说明有更少的未迁移样本被判断为迁移，同时查全率有了一定的降低，代表迁移样本被全部准确判断的可能性降低了。综合来看是(v_{He})降低了所有样本被判断为迁移样本的条件，因此，可以认为氦原子总速度大小会对判断氦原子迁移有一定的影响，在提高了迁移样本被找出的概率的同时也会使得更多的未迁移样本被判断为迁移样本。

去除 3 个氦原子在三个方向上的速度分量(v_x, v_y, v_z)得到了更低的查准率和更高的查全率，这意味着三个速度分量提高了判断为迁移样本的条件。具有更严格条件的三速度分量样本才能被判断为迁移样本，因此，可以认为三速度分量对判断氦是否发生迁移是较为重要的。

去除 3 个氦原子的速度方向夹角($\theta_x, \theta_y, \theta_z$)，模型的预测能力未发生改变。产生这种情况的原因可能是因为其余特征已经包含了该部分特征信息，因此速度方向夹角特征对该模型来说可以认为是非必要的。

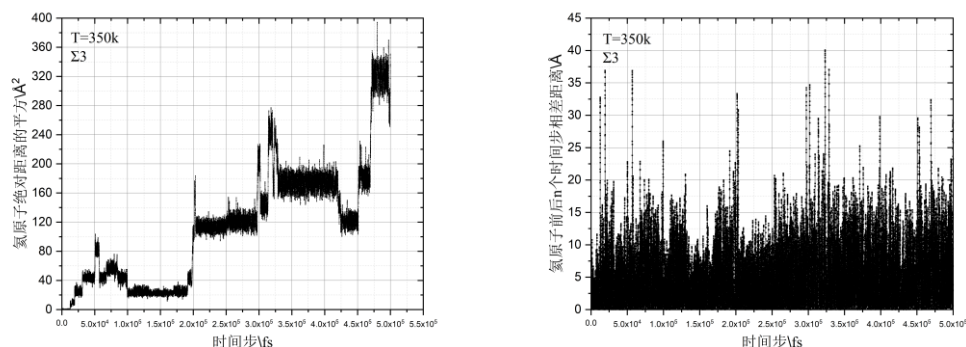
去除 6 个氦原子的铁氦结构参数($CS_2, CS_4, CS_6, CS_8, CS_{10}, CS_{12}$)带来的结果是查准率未变，查全率下降。这说明该参数在提高判断为迁移样本的条件的同时，不会让更多的未迁移样本判断为迁移样本。同时查全率的改变值相对于其他参数来说变动是最大的，这说明铁氦结构参数对发生迁移的样本的影响要更大一些，仅依靠氦原子的运动状态判断氦原子迁移是不可靠的。

仅有 6 个氦原子的铁氦结构参数($CS_2, CS_4, CS_6, CS_8, CS_{10}, CS_{12}$)带来的结果是查准率、查全率和 F1 分数均有大幅下降的结果，说明没有氦原子的运动状态是难以判断间隙氦是否迁移的。

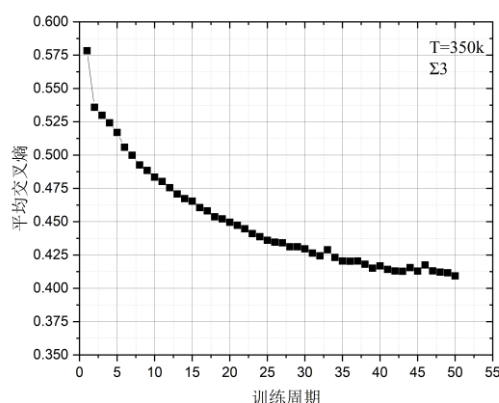
综合来看，完整晶粒中氦原子的速度特征是判断间隙氦在 350K $\alpha\text{-Fe}$ 晶粒中是否迁移的重要因素，同时铁氦结构参数也在一定程度上能提高整体的判断效果。任意去除其中的一种速度特征对模型的影响不大。

(b) $\alpha\text{-Fe}\langle 110 \rangle$ 倾斜轴 $\Sigma 3\{112\}$ 晶界，350K，晶界能为 260mJ/m^2 ，演化时间 1ns

同样的方法，在该晶界中进行氦原子迁移演化了 1ns，每 10fs 输出了一帧氦原子迁移特征，一共有 10 万帧。氦原子均方位移和一个时刻前后 20 个时刻的氦原子相对于初始位置的距离和的绝对差值如图 4.6 所示(这里仅展示 5ns)：

图 4.6 $\Sigma 3$ 氢原子迁移情况

根据上图结果，结合 dump 文件的直观结果，最终挑选了具有可靠分类结果的 6200 个迁移样本和 6200 个未迁移样本。同样通过每个样本自身时刻之前的连续的 40 个样本的氢原子迁移特征加上本身时刻的氢原子迁移特征共同组成该时刻 LSTM 网络的输入特征。打乱后随机选择 9920 个样本作为训练样本，2480 个样本作为测试样本。训练集上的平均交叉熵随训练周期的变化如下图 4.7 所示：

图 4.7 $\Sigma 3$ 模型训练平均交叉熵随训练周期的变化

从上图可以看到，达到 50 个训练周期时，该模型在训练集上已经基本收敛，再对该模型的预测效果进行评价，得到结果为测试集中 1270 个迁移样本中有 1105 个样本被正确判断为发生迁移，1210 个未迁移样本中有 970 个样本被判断为未发生迁移。具体的评价指标如表 4.3 所示：

表 4.3 LSTM 模型在训练集上评价指标

输入特征	查准率	查全率	F1 分数
------	-----	-----	-------

四类特征	0.82	0.87	0.84
无(v_{He})	0.88	0.83	0.835
无(v_x, v_y, v_z)	0.87	0.82	0.84
无($\theta_x, \theta_y, \theta_z$)	0.81	0.88	0.84
无铁氢结构参数	0.80	0.83	0.82
仅铁氢结构参数	0.77	0.81	0.79

上表可得在 $\Sigma 3$ 晶界模型中通过 LSTM 模型也能够对间隙氢原子是否迁移进行判断。下面同样的从四个特征来具体分析。

去除氢原子总速度大小(v_{He})，模型查准率大幅上升，查全率大幅下降。与完整晶体模型中的表现趋势相同，但在 $\Sigma 3$ 晶界模型中改变幅度更为剧烈，两模型的温度是相同，这说明总速度大小在晶界模型中更大地减低了所有样本被判断为迁移样本的条件。在一定程度上这代表了在 $\Sigma 3$ 晶界更多的未迁移样本和迁移样本的总速度大小接近。

去除 3 个氢原子在三个方向上的速度分量(v_x, v_y, v_z)，模型的查准率大幅提高、查全率大幅下降。这与完整晶体模型的情况相反，完整晶体中三速度方向对样本判断为迁移样本的条件是增加的，而在该晶界模型中，却降低了条件。这代表在完整晶体中样本迁移对速度分量的要求更为苛刻，而在晶界模型中该条件被弱化了。

去除 3 个氢原子的速度方向夹角($\theta_x, \theta_y, \theta_z$)，模型查准率略有降低，模型查全率略有提高。这也与完整晶体中的情况不一致，这说明了在 $\Sigma 3$ 晶界模型中速度方向夹角提高了被判断为迁移样本的条件，而该参数也对氢原子的迁移具有比完整晶体模型中更大的影响。

去除 6 个氢原子的铁氢结构参数($\text{CS}_2, \text{CS}_4, \text{CS}_6, \text{CS}_8, \text{CS}_{10}, \text{CS}_{12}$)带来的结果是查准率、查全率均下降。在完整晶体中该参数只对发生迁移的样本产生了较大的影响，而在 $\Sigma 3$ 晶界模型中该参数对所有样本的正确判断均具有很重要的价值。没有铁氢结构参数将导致模型的判断能力整体下降。

综合来看，在 350K $\alpha\text{-Fe}$ 的 $\Sigma 3$ 晶界中，相较于完整晶粒，氢速度分量对判断氢原子迁移的影响减弱，而氢速度方向夹角和中心对称参数的影响上升，仅有铁氢对称参数仍会导致所有参数的下降，这说明间隙氢运动特征仍旧占据重要的作用。

(c) $\alpha\text{-Fe}\langle 110 \rangle$ 倾斜轴 $\Sigma 11\{113\}$ 晶界，500K，晶界能为 1113mJ/m^2 ，演化时间 1ns
同样的在 $\Sigma 11$ 晶界模型中也做了相应的间隙氢原子迁移模拟。但由于在 350K

时,该晶界模型中的间隙氦仅有少数发生了迁移,因此为了保证迁移样本的丰富性,在本例中将温度提高到了 500K,具体的氦原子均方位移和前后时刻距离和的绝对差值随演化时间的变化如下图 4.8 所示(这里仅展示 5ns):

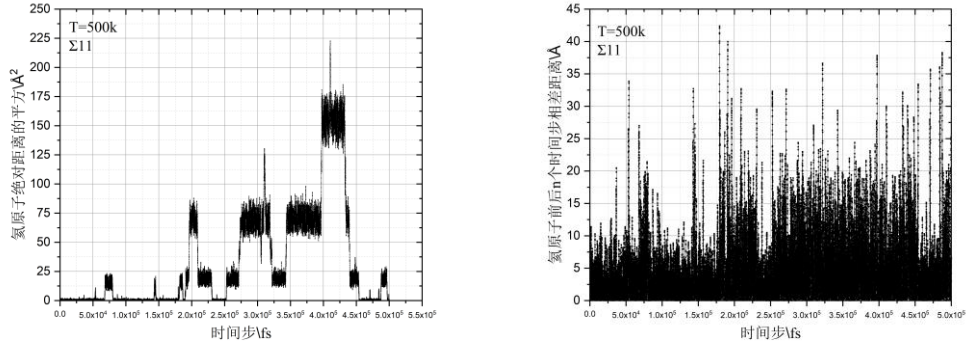


图 4.8 $\Sigma 11$ 氦原子迁移情况

根据上图结果,挑选了 4000 个迁移样本和 4000 个未迁移样本。同样通过每个样本自身时刻之前的连续的 40 个样本的氦原子迁移特征加上本身时刻的氦原子迁移特征共同组成该时刻 LSTM 网络的输入特征。打乱后随机选择 6400 个样本作为训练样本,1600 个样本作为测试样本。随后利用 LSTM 模型对该训练集做了训练,具体的训练集上平均交叉熵随训练周期的变化如下图 4.9 所示:

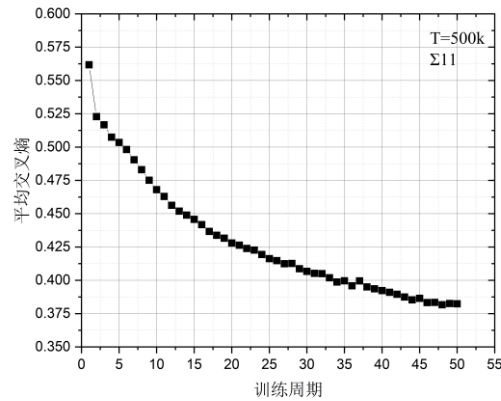


图 4.9 $\Sigma 11$ 模型训练平均交叉熵随训练周期的变化

同样的,从上图可知在 50 的训练周期时,模型已基本收敛。该模型在测试集的表现 832 个发生迁移样本中 707 个样本被找到,而 768 个未迁移样本中有 155 个样本判断为发生迁移,其余 613 个被正确判断。

表 4.4 LSTM 模型在训练集上评价指标

输入特征	查准率	查全率	F1 分数
四类特征	0.82	0.85	0.83
无(v_{He})	0.88	0.88	0.88
无(v_x, v_y, v_z)	0.86	0.89	0.87
无($\theta_x, \theta_y, \theta_z$)	0.84	0.87	0.85
无铁氦结构参数	0.87	0.77	0.82
仅铁氦结构参数	0.83	0.69	0.75

从上表中同样可以得到 LSTM 模型可以对 $\Sigma 11$ 晶界模型中的间隙氦原子进行迁移情况的判断,准确率为 80%的水平。接下来对四种特征逐一进行比较分析。

去除氦原子总速度大小(v_{He}),模型的查准率查全率均得到了较大的提升,出现了与前两模型均不同的现象,这说明了在 $\Sigma 11$ 晶界模型中总速度大小已然对判断氦是否迁移有了一定的干扰因素,更多的迁移样本和未迁移样本具有相同的总速度大小,因此总速度大小对是否迁移的判断作用则减弱了。

去除 3 个氦原子在三个方向上的速度分量(v_x, v_y, v_z),模型的查准率查全率变化与去除(v_{He})类似,同样得到了较大的提升。这说明了速度分量也随着温度效应的增加逐渐难以判断氦原子是否迁移了。这在前两模型中均未有相同的表现,可见温度效应对模型的影响是十分巨大的。

去除 3 个氦原子的速度方向夹角($\theta_x, \theta_y, \theta_z$),模型的查准率查全率同样有了提升,但提升的幅度不如(v_{He})和(v_x, v_y, v_z)。可见随着温度的增大,在 $\Sigma 11$ 晶界模型中,所有代表温度效应的氦迁移特征的分类作用均开始下降。

去除 6 个氦原子的铁氦结构参数($CS_2, CS_4, CS_6, CS_8, CS_{10}, CS_{12}$)带来的结果是查准率上升,查全率极具下降,这说明了铁氦结构参数降低了被判断为迁移间隙氦原子的条件。以铁氦结构参数作为迁移特征,将有更多的样本被判断为迁移氦原子。

可以看到四种减少特征的方式均使得模型的查准率上升了,这说明了在该模型中过多的特征导致了模型受到了较多的干扰,这也与温度上升未迁移氦原子发生更剧烈的热运动有关。同时,没有间隙氦速度特征仍旧会带来模型的整体下滑。

(d) $\alpha\text{-Fe}\langle 110 \rangle$ 倾斜轴 $\Sigma 11\{113\}$ 晶界, 350K, 晶界能为 1113mJ/m², 演化时间 1ns

同样的方法在 4000 个迁移样本和 4000 个未迁移样本上,随机抽取 6400 个作为训练集,剩下 1600 个作为测试集。分析各个微观特征对模型造成的影响,如表 4.5 所示。

表 4.5 LSTM 模型在训练集上评价指标

输入特征	查准率	查全率	F1 分数
四类特征	0.83	0.80	0.81
无(v_{He})	0.84	0.78	0.81
无(v_x, v_y, v_z)	0.79	0.85	0.82
无($\theta_x, \theta_y, \theta_z$)	0.78	0.75	0.78
无铁氦结构参数	0.74	0.71	0.72
仅铁氦结构参数	0.81	0.68	0.74

可以看到，间隙氦运动特征对判断间隙氦是否迁移仍有重要的作用。但无铁氦对称参数带来的模型效果下降超过了无间隙氦运动特征带来的下降。这说明在该模型中结构信息的重要程度甚至超过了运动特征。

(e) $\alpha\text{-Fe}\langle 110 \rangle$ 倾斜轴 $\Sigma 11\{332\}$ 晶界，500K，晶界能为 1019mJ/m²，演化时间 1ns

同样的方法在 2000 个迁移样本和 2000 个未迁移样本上，随机抽取 3200 个作为训练集，剩下 800 个作为测试集。分析各个微观特征对模型造成的影响，如表 4.6 所示。

表 4.6 LSTM 模型在训练集上评价指标

输入特征	查准率	查全率	F1 分数
四类特征	0.91	0.84	0.87
无(v_{He})	0.92	0.80	0.85
无(v_x, v_y, v_z)	0.83	0.82	0.82
无($\theta_x, \theta_y, \theta_z$)	0.93	0.80	0.86
无铁氦结构参数	0.90	0.80	0.86
仅铁氦结构参数	0.74	0.81	0.77

可以看到，在高温环境下，速度特征对判断间隙氦是否迁移是有重要的作用的。

为了更好地挖掘各微观新宿对间隙氦迁移的影响，以 F1 分数改变百分比作为评价标准，综合上面五种晶界模型，共同分析各因素的影响，如下图所示：

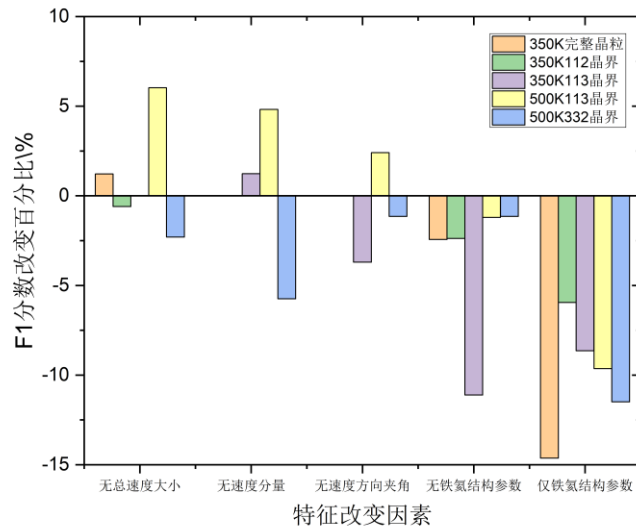


图 4.10 F1 分数随各因素缺失的变化图

观察可见，无铁氢结构参数时所有模型效果均下降。仅有铁氢结构参数时模型效果下降幅度更大。个别晶界上出现 F1 分数增长，仅有铁氢结构参数的时候模型效果均大幅下降。同一温度(350K)下，模型在{113}晶界(1113mJ/m²)上对速度方向夹角和铁氢结构参数的敏感度远高于完整晶粒和{112}晶界(260mJ/m²)。{113}晶界在 350K 下铁氢结构参数的重要性高过了间隙氢运动特征，而在 500K 时，间隙氢运动特征相较于铁氢结构参数的重要性更高了。

4.5 本章小结

通过 LSTM 算法对间隙氢原子在 α -Fe 晶界中的迁移情况进行了判断。该模型判断间隙氢原子是否迁移的查准率、查全率和 F1 分数都能达到 0.8 的水平，因此，这说明通过在 4.2 章节中构造的四种间隙氢原子迁移特征是能够较大得满足判断氢原子是否迁移的条件。同时，分析了这四类参数的对模型的影响，在 350K 完整晶体中速度分量大小有着重要的影响因素，而速度方向夹角则可以被其余特征替代。在 350K $\Sigma 3$ 晶界模型中，氢速度方向夹角和铁氢结构参数对判断氢原子是否迁移的影响较完整晶体是上升的。在 500K $\Sigma 11$ 晶界模型中，由于温度的上升，原子热运动加剧，速度相关特征均对模型判断产生了一定的干扰作用，而铁氢参数使得更多的迁移样本被判断出来了。间隙氢所处环境(铁氢结构参数)对判断是否迁移具有重要的作用。间隙氢的速度特征较环境特征对判断迁移更为重要。速度特征的冗余可能对模型造成负面影响，但不能完全没有速度特征。

第 5 章 总结与展望

5.1 总结

全文通过两个任务分别构建了晶界微观结构与晶界能间的关联以及氢原子微观信息与氢原子迁移情况之间的联系。为下一步定量分析微观结构和宏观性质间的连接提供了基础。

在预测晶界能的任务中得到了以下的结论：在 TextCNN 模型的基础上，相比于径向分布函数和中心对称参数，根据晶界原子密度信息和局部紊乱程度构造的积累中心对称参数能够与晶界能间建立更紧密的关联。该参数作为晶界特征利用 TextCNN 模型在 α -Fe 的 $\langle 100 \rangle$ 和 $\langle 110 \rangle$ 倾斜轴的对称倾斜晶界上预测晶界能的平均相对误差能达到 1.74%。为进一步研究 RAFM 钢晶界工程提供了支撑。

通过判断间隙氢在晶界中是否迁移的任务得到的结论有：对于 α -Fe 完整晶粒中的氢原子在 350K 演化 1ns 的模型，氢原子总速度大小、速度分量和铁氢结构参数对判断间隙氢是否迁移有较大的影响而氢原子的速度方向夹角在有另三种特征参数的情况下显得不那么重要。对于 α -Fe 的 $\Sigma 3$ 晶界中的氢原子在 350K 下演化 1ns 的模型，四种特征对判断间隙氢迁移情况有不同的影响，其中铁氢结构参数的影响相较于晶粒中有了一定的提升。对于 α -Fe 的 $\Sigma 11$ 晶界中的氢原子在 500K 下演化 1ns 的模型，氢原子总速度大小、速度分量和速度方向夹角对判断间隙氢是否迁移在有其余特征时均有一定的干扰影响，而铁氢结构参数提高了判断间隙氢迁移的可靠性。通过分析这些参数与间隙氢迁移情况之间的关联为下一阶段定量分析微观氢原子信息与宏观氢状态间的连接打下了基础。

5.2 展望

本文主要通过深度学习方法在一定程度上建立了微观原子信息与宏观材料性质之间的关联，但这种关联性还不能通过定量的表现出来，因此下一步希望在目前研究的基础上定量地去挖掘晶界微观结构与晶界宏观性能间的关联以及氢原子的微观信息与宏观氢泡形成之间的关联。为聚变堆结构材料的研究添砖加瓦。

参考文献

- [1] 杨文斗. 反应堆材料学[M]. 北京:原子能出版社, 2000.
- [2] 王旭. 奥氏体钢和铁素体-马氏体钢的离子辐照效应[D]. 2015.
- [3] Bloom E F, Conn R W, Davis J W, et al. Low Activation Materials for Fusion Applications[J]. J Nucl Mater, 1983, 123:17.
- [4] Klueh R I, Gelles D S, Jitsukawa S, et al. Ferritic/Martensitic Steels-Overview of Recent Results[J]. J Nucl Mater, 2002, 307-312:455.
- [5] 黎兴刚, 燕青芝, 葛昌纯等. 低活化铁素体/马氏体钢的研究进展[J]. 钢铁研究学报, 2009, 21:6.
- [6] Gilbert MR, Dudarev SL, Zheng S, et al. An integrated model for materials in a fusion power plant: transmutation, gas production, and helium embrittlement under neutron irradiation [J]. Nuclear Fusion, 2012, 52:1-12.
- [7] 余龙, 陈李嵘, 隋昊男等. 金属材料力学性能的辐照氦效应研究进展 [J]. 固体力学学报, 2020, 41(6).
- [8] Trinkaus H, Singh B N. Helium accumulation in metals during irradiation -where do we stand? [J]. Journal of Nuclear Materials, 2003, 323(2-3):229-242.
- [9] 杨磊, 王苍龙, 祁美灵. 中国 CIADS 发展及结构材料辐照多尺度模拟方法研究 [C]. 中国力学大会, 2015.
- [10] Fu C C, Willaime F. Ab initio study of helium in α -Fe: Dissolution, migration, and clustering with vacancies [J]. Physical Review B, 2005, 72(6): 064117.
- [11] 严六明, 朱素华. 分子动力学模拟的理论与实践 [M]. 北京: 科学出版社, 2013.
- [12] DE Makarov, PK Hansma, H Metiu. Kinetic Monte Carlo simulation of titin unfolding [J]. Journal of Chemical Physics, 2001, 114:21.
- [13] D. Y. Sun, M. Asta and J. J. Hoyt. Crystal-melt interfacial free energies in hcp metals: A molecular dynamics study of Mg [J]. Phys. Rev. B, 2004, 69.
- [14] D. Wolf. Correlation between the energy and structure of grain boundaries in b.c.c. metals. II. Symmetrical tilt boundaries [J]. Philos. Mag. A, 1990, 62:447.
- [15] H. Nakashima and M. Takeuchi. Grain boundary energy and structure of α -Fe $\langle 110 \rangle$ symmetric tilt boundary [J]. Tetsu-to-Hagane, 2000, 86:357.
- [16] T. Nagano and M. Enomoto. Calculation of the interfacial energies between α and γ iron and equilibrium particle shape [J]. Mater: Metall. Trans. A, 2006, 37A.
- [17] Sutton A. P., Balluffi R. W. 晶体材料中的界面 [M]. 北京: 高等教育出版社, 2016.

- [18] Bollmann W. Crystal defects and crystalline interfaces [M]. Berlin: Springer-Verlag, 1970.
- [19] Jingyi Shi, Fei Gao, Lei Peng, Minyou Ye. Molecular Dynamics Study: Effects of He Bubble and Cr precipitate on Tensile Deformation of Grain Boundaries in α -Fe [J]. IEEE Trans, 2017, 45:289-293.
- [20] 张明亮, 杨亮等. 基于修正球形双晶模型的金属 Al 晶界能分子动力学计算 [J]. 中国有色金属学报, 2015, 25(11).
- [21] READ W T, SHOCKLEY W. Dislocation models of crystal grain boundaries[J]. Physical Review, 1950, 78(3): 275–289.
- [22] HASSON G C, GOUX C. Interfacial energies of tilt boundaries in aluminium: Experimental and theoretical determination[J]. Scripta Metallurgica, 1971, 5(10): 889–894.
- [23] Jingyi Shi, Fei Gao, Lei Peng. Effects on Grain Boundary Strength in Fe: A Molecular Dynamics Simulation [C]. SOFE, 2016.
- [24] Joshua A. Gomberg. Extracting knowledge from molecular mechanics simulations of grain boundaries using machine learning [J]. Acta Mater, 2017, 133:100–108.
- [25] S.R. Kalidindi, J.A. Gomberg, Z.T. Trautt, C.A. Becker. Application of data science tools to quantify and distinguish between structures and models in molecular dynamics datasets [J]. Nanotechnology, 2015, 26(34).
- [26] X. Dong, D.L. McDowell, S.R. Kalidindi, K.I. Jacob. Dependence of mechanical properties on crystal orientation of semi-crystalline polyethylene structures [J]. Polymer, 2014, 55(16).
- [27] Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning [M]. 北京:人民邮电出版社, 2017.
- [28] Andreas Mardt, Luca Pasquali. VAMPnets: Deep learning of molecular kinetics [J]. Nature Communications, 2018, 9(5).
- [29] Brain L. DeCost, Bo Lei. High Throughput Quantitative Metallography for Complex Microstructures Using Deep Learning: A Case Study in Ultrahigh Carbon Steel [J]. Microscopy and Microanalysis, 2019, 25:21-29.
- [30] 闫海, 邓忠民. 基于深度学习的短纤维增强聚氨酯复合材料性能预测 [J]. 复合材料学报, 2019, 36(6): 1413-1420.
- [31] 李航. 统计学习方法[M]. 北京:清华大学出版社, 2012.
- [32] Gareth James, Daniela Witten. An introduction to statistical learning [M]. New York: Springer, 2017.

- [33] W S McCulloch, W Pitts. A logical calculus of the ideas immanent in nervous activity [J]. Bull. Math. Biophys., 1943, 5:115-137.
- [34] F Rosenblatt. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms [M]. Washing D.C., Spartan, 1961.
- [35] J L McClelland, D E Rumelhart. Parallel distributed processing : exploration in the microstructure of cognition [M]. Boston, MIT Press, 1986.
- [36] LeCun Y, Bengio Y. Convolutional Networks for Images, Speech, and Time-Series [J]. The Handbook of Brain Theory and Neural Networks, 1995.
- [37] ZC Lipton, J Berkowitz, C Elkan. A Critical Review of Recurrent Neural Networks for Sequence Learning[J]. arXiv, 2015, 1506.
- [38] Plimpton S, Crozier P, Thompson A. LAMMPS large scale atomic /molecular massively parallel simulator [J]. J. Appl. Phys., 2015, 20: 4740.
- [39] Pumphrey PH, Malis TF and Gleiter H. Rigid body translations at grain boundaries[J]. Philosophical Magazine, 1976, 34:227-233.
- [40] Ackland GJ, Mendelev MI, Srolovitz DJ. Development of an interatomic potential for phosphorus impurities in alpha-iron [J]. Journal of Physics: Condensed Matter, 2004, 16: S2629-S2642.
- [41] Yasushi SHIBUTA, Shinya TAKAMOTO and Toshio SUZUKI. A Molecular Dynamics Study of the Energy and Structure of the Symmetric Tilt Boundary of Iron [J]. ISIJ International, 2008, 48:1582–1591.
- [42] D. Wolf. Correlation between the energy and structure of grain boundaries in b.c.c. metals. II. Symmetrical tilt boundaries. Philos. Mag. A, 1990, 62:447.
- [43] C.L.Kelchner, S.J.Plimpton, J.C.Hamilton. Dislocation nucleation and defect structure during surface indentation [J]. Physical Review B, 1998, 58:11085.
- [44] Harrington P. Machine learning in action [M]. 北京:人民邮电出版社, 2013.
- [45] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2828-2826.
- [46] K He, X Zhang, S Ren, J Sun. Deep Residual Learning for Image Recognition [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [47] Sahil Babbar. Review - Mastering the game of Go with deep neural networks and tree search [J]. IIIT-Delhi, 2017.
- [48] Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines [C]. Proceedings of the 27th International Conference on Machine Learning, 2010: 807-

814.

[49] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors) [J]. *The Annals of Statistics*, 2000, 28(2):337-407.

[50] Murray N, Perronnin F. Generalized max pooling [C]. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 2473-2480.

[51] Boureau Y L, Ponce J, LeCun Y. A theoretical analysis of feature pooling in visual recognition [C]. *Proceedings of the 27th International Conference on Machine Learning*, 2010:111-118.

[52] Liu W, Wen Y, Yu Z, et al. Large-margin softmax loss for convolutional neural networks [C]. *Proceedings of the 33rd International Conference on Machine Learning*, 2016:507-516.

[53] Yoon Kim. Convolutional Neural Networks for Sentence Classification [J]. *arXiv preprint*, 2014, arXiv:1408.5882.

[54] S Ioffe, C Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [J]. *arXiv*, 2015, arXiv:1502.

[55] J Duchi, E Hazan, Y Singer. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization [J]. *Journal of Machine Learning Research*, 2011.

[56] KINGMA D, BA J. Adam: A method for stochastic optimization[C] // *The 3rd International Conference on Learning Representations*. CA: ICLR Committee, 2015.

[57] Inoue H. Data Augmentation by Pairing Samples for Images Classification[J]. *ICLR*, 2018.

[58] B. Schölkopf, K. K. Sung, C. J. C. Burges. Comparing support vector machines with Gaussian kernels to radial basis function classifiers [J]. *IEEE Trans: Signal Process*, 1997, 45:2758–2765.

[59] Wolpert, D.H., Macready, W.G. No Free Lunch for optimization [J]. *IEEE Transactions on Evolutionary Computation*, 1997, 1(1):67-83.

[60] Barbes R S. Embrittlement of stainless steels and nickel-based alloys at high temperature induced by neutron radiation [J]. *Nature*, 1956, 206(4991):1307-1310.

[61] Ullmaier H. Helium in metals [J]. *Radiat. Eff.*, 1983, 78(1-4):1-10.

[62] Morishita K, Sugano R, Wirth B D. MD and KMC modeling of the growth and shrinkage mechanisms of helium-vacancy clusters in Fe [J]. *J. Nucl. Mater.*, 2003, 323(2-3):243-250.

[63] Gao F, Heinisch H, Kurtz R J. Diffusion of He interstitials in grain boundaries in

α -Fe [J]. J. Nucl. Mater., 2006, 351(1-3):133-140.

[64] Gao F, Heinisch H, Kurtz R J. Migration of vacancies, He interstitials and He-vacancy clusters at grain boundaries in α -Fe [J]. J. Nucl. Mater., 2009, 386-388:390-394.

[65] S Hochreiter. LSTM can solve hard long term lag problems [J]. Neural Information Processing Systems Nips, 1997.

[66] Liu W, Wen Y, Yu Z, et al. Large-margin softmax loss for convolutional neural networks [C]. Proceedings of The 33rd International Conference on Machine Learning, 2016: 507-516.

致 谢

一转眼三年研究生生涯已经接近尾声，好像自己昨天才刚参加完考研复试。这三年的酸甜苦辣历历在目，从一开始的懵懵懂懂到现在的目标明确，也算是为我这 10 多年来的求学生涯画上了最圆满的句号。

这两年，疫情袭来，生命的脆弱赤裸裸地展现在人们眼前，我告诫自己，要努力且坚定地过好每一天。

感谢自己能生活在这样一个充满希望的时代，也感谢自己的父母将我养育长大。我的母亲姚冬萍教会了我坚强，她包容我的一切，给予了我无私的爱。

感谢我的导师彭蕾老师和实验室的各位兄弟姐妹，让我能在科大拥有一个舒适的科研环境。

感谢我的女朋友汪辰，她的阳光和充满勇气经常感染着我，不过我希望她能适度追星。

感谢我爱听的那些歌，能给我带来平静。

最后，感谢我自己，能一直保持最初的真心。

陈 村
2021.5