

基于多源数据融合的半导体晶片 CMP 抛光材料去除率预测

方维, 王宇宇, 宋志龙, 吕冰海, 赵文宏*

(浙江工业大学 机械工程学院, 杭州 310023)

摘要: **目的** 对半导体晶片抛光过程中的工艺参数、耗材使用量、抛光垫状态参数等多源数据预处理后进行数据融合, 建立材料去除率 (MRR) 预测模型, 为实现半导体晶片抛光加工工艺的决策和处理奠定基础。**方法** 研究晶片抛光加工中的数据特点及数据融合需求, 提取数据集中每个晶片加工过程中的统计特征并生成新数据集, 同时引入邻域特征以应对晶片加工过程中动态因素对材料去除率的影响。提出基于深度自动编码器的多源数据融合及材料去除率预测方法。设计深度自动编码器参数, 优化深度自动编码器的损失函数从而增强深度自动编码器对强相关性特征变量的重建。基于深度自动编码器进行多源传感器信号融合, 降低数据维度。使用超参数搜索算法优化 BP 神经网络超参数, 利用 BP 神经网络方法将融合后的数据进行半导体晶片抛光过程中的材料去除率预测。**结果** 采用 PHM2016 数据集对模型进行验证, 均方误差 MSE 达到 7.862, 相关性 R^2 达到 91.2%。**结论** 基于多源数据的融合模型能有效预测 MRR, 可以对半导体晶片 CMP 工艺过程的智能决策与控制起到良好的辅助作用。

关键词: 化学机械抛光; 材料去除率; 数据融合; 深度自动编码器; BP 神经网络预测

中图分类号: TN305.2 **文献标志码:** A **文章编号:** 1001-3660(2024)02-0150-08

DOI: 10.16490/j.cnki.issn.1001-3660.2024.02.014

Prediction of CMP Polishing Material Removal Rate of Semiconductor Wafers Based on Multi-source Data Fusion

FANG Wei, WANG Yuyu, SONG Zhilong, LYU Binghai, ZHAO Wenhong*

(College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou 310023, China)

ABSTRACT: The multi-source data, such as process parameters, consumption of consumables and polishing pad state parameters, were preprocessed for multi-source data fusion, and the material removal rate (MRR) prediction model was established to realize the decision of semiconductor wafer polishing process. The data characteristics and data fusion requirements of multiple sources of data in wafer polishing were analyzed. The statistical features of each wafer in the original data set were extracted to create a new data set. In order to address the impact of dynamic factors in wafer processing on the material removal rate, the neighborhood features were introduced. There were multiple sources and multiple modes of data in polishing of semiconductor wafers, such as the contact pressure between the wafer and the vehicle, the flow rate change of the

收稿日期: 2023-01-01; 修订日期: 2023-07-05

Received: 2023-01-01; Revised: 2023-07-05

基金项目: 国家自然科学基金 (U20A20293); 浙江省自然科学基金 (LD22E050010)

Fund: National Natural Science Foundation of China (U20A20293); Natural Science Foundation of Zhejiang Province (LD22E050010)

引文格式: 方维, 王宇宇, 宋志龙, 等. 基于多源数据融合的半导体晶片 CMP 抛光材料去除率预测[J]. 表面技术, 2024, 53(2): 150-157.
FANG Wei, WANG Yuyu, SONG Zhilong, et al. Prediction of CMP Polishing Material Removal Rate of Semiconductor Wafers Based on Multi-source Data Fusion[J]. Surface Technology, 2024, 53(2): 150-157.

*通信作者 (Corresponding author)

polishing slurry, and the rotational speed of the polishing disc. There was a certain correlation between the data of these modalities, each modality could provide specific information for the other modal data. Processing the data of different modalities in the same way or simply concatenating all modal features could not guarantee the effectiveness of subsequent regression prediction. An efficient method to combine and reduce the dimensionality of these multiple sources and multiple modes of data was required. The material removal rate prediction method with deep automatic encoder for multi-source data fusion was proposed: the deep autoencoder parameters was designed and the loss function of the deep autoencoder was optimized to enhance the reconstruction of strongly correlated characteristic variables by the deep autoencoder. Considering the degree of data completeness of samples, completeness weights were introduced to reduce the impact of incomplete data on model training. Multi-source sensor signal fusion and data dimensionality reduction based on deep autoencoder were performed, and the fused data were fed into a BP neural network for MRR prediction during semiconductor wafer polishing (The hyperparameters of the BP neural network were optimized by the Hyperband hyperparameter search algorithm). The model was verified in the PHM2016 data for comparative analysis with other models (linear regression model, KNN regressor model and BP neural network model). The model in this study was based on an deep auto-encoder to achieve feature fusion and dimensionality reduction, and the processed data were employed in a BP neural network to fit the material removal rate of semiconductor wafers. Compared with the linear regression model (the mean square error MSE reached 31.186), the KNN regressor model (the mean square error MSE reached 26.17 and the correlation R^2 reached 54.3%) and the BP neural network model (the mean square error MSE reached 11.445 and the correlation R^2 reached 82.4%), the model in this study had a prediction value closer to the actual value in the three scenarios, with the mean square error MSE reached 7.862 and the correlation R^2 reached 91.2%. The model based on multi-source data fusion can effectively predict MRR, which plays a benign auxiliary role in intelligent decision-making and control of the CMP process of semiconductor wafers.

KEY WORDS: CMP; MRR; data fusion; deep auto-encoder; BP neural network prediction

化学机械抛光 (Chemical Mechanical Polishing, CMP) 是半导体制造行业中实现电子材料表面平坦化的一项基础技术^[1], 最早由 Monsanto 于 1965 年提出, 具有加工成本低及加工方法简单的优点, 现已成为半导体加工领域中的核心技术^[2]。对于半导体晶片材料去除率、表面粗糙度等参数, 通常在晶片完成抛光并清洁后进行检测, 很难在加工过程中实现实时检测, 这会导致抛光加工工艺决策时间延长。可以在工件表面还未产生缺陷前对工艺参数进行预调整, 从而减少确定抛光工艺的工作量^[3]。

目前, 半导体晶片 CMP 材料去除率建模预测方法有物理方法和数据驱动方法 2 种。基于物理模型的方法是通过目标的物理特性建立数学模型, 从而反映目标系统运行或退化的机理。Xu 等^[4]基于固-固接触力学、流体动力学、化学反应动力学和磨损理论, 建立了完整的晶片材料去除率模型框架, 仿真结果与实验数据吻合较好, 表明建立的计算模型是可行的。杜家熙等^[5]基于硅片化学机械抛光过程中的摩擦磨损行为, 建立了硅片材料去除率模型, 结果表明, 在晶片 CMP 过程中主要的机械作用由磨料产生, MRR 主要是由机械作用和化学作用相互作用而产生的。蒋建忠等^[6]提出在典型的机械化学抛光过程中芯片表面材料的去除应为表面非晶层物质黏性流动所致的新观点, 他们建立了一种新的表征 CMP 过程中材料去除速率的数学模型, 结果表明, 该模型的理论预测值

与实验测定值十分吻合。Liu 等^[7]在熔融二氧化硅 CMP 过程中, 考虑了机械作用和化学腐蚀对材料去除的影响, 建立了新的材料去除率模型, 实验结果和预测结果表明, CMP 材料去除率与抛光压力和抛光液浓度之间存在非线性关系, 但受到抛光垫状态、材料性能、加工环境的耦合作用, 工艺参数、耗材使用量、抛光垫状态参数等多源数据与材料去除率的非线性映射关系难以通过物理模型写出具体数学表达式。

基于数据驱动模型能提取与预测目标相关的数据, 完成数据融合、自主分析预测或诊断。与物理驱动模型相比, 数据驱动模型具有能通过抽检晶片的 MRR 来矫正模型预测值的优势。越来越多的人开始使用数据驱动的模型来预测 CMP 中的 MRR^[8]。Di 等^[9]在 CMP 建模中结合了物理机制和最近邻的影响, 提出了一种改进的 VM 方法。通过该方法获得了晶片加工过程中详细的动态变化情况, 他们根据这些信息动态修正了当前加工晶片的 MRR 预测值, 并通过 PHM 2016 数据集对该方法进行了评价, 得到的均方误差为 7.07。Jia 等^[10]基于 GMDH 类型的神经网络提出了一种新的具有适应性的 VM 模型, 在模型中使用了时间邻域特征和耗材使用量邻域特征以应对加工过程中的动态变化, 模型的均方误差为 6.62。Feng 等^[11]提出了一种具有动态特征选取和在线模型学习的 VM 模型, 该模型与 DL (一个具有代表性的样本子集) 基于从数据库中随机选择的少量样本进行初始

化,在监测中迭代更新数据库,追踪加工过程中由耗材等因素引起的 MRR 动态变化。Cai 等^[12]提出了一种高斯回归过程 (GPR) 模型,动态预测了 CMP 中的 MRR,该模型在 PHM2016 数据集中的均方差评价指标达到 3.4。

随着机器学习的广泛应用^[13],多源数据融合技术能够更好地分析由传感器采集的具有多样性和复杂性的数据^[14],在工艺流程的动态智能决策和先进加工制造中,深度学习方法发挥着重要的作用,可以预测工件的材料去除率,同时能够实现对加工过程的实时多维感知检测。本文利用深度学习方式提取多源数据中的有效信息^[15-17],提出基于深度自动编码器 (Deep AutoEncoder, DAE) 进行多源传感器信号融合、降维,利用 BP 神经网络 (Back Propagation Neural Network, BPNN) 深度学习方法对融合后的数据进行晶片材料去除率的预测,最终得到基于 DAE 的 MRR 预测模型,预测不同工艺参数下的材料去除率,使加工工艺得到决策和处理。

1 数据预处理及特征提取

1.1 数据预处理

选择 PHM2016^[18]比赛中的数据集作为本文数据集。PHM2016 数据提供了状态监测数据和测量结果数据。其中状态监测数据是每个抛光过程中各个场所工况 (分为精抛、粗抛) 及开关测量 (例如抛光液开关是否打开) 结果的时间序列形式数据,包括耗材使用量 (7-12)、压力 (13-18)、抛光液流量 (19-21)、转速 (21-24)、修整液状态 (25) 等共计 25 列,其具体名称如表 1 所示。测量结果数据是针对每个抛光过程,人工测量计算得到经抛光加工后晶片的平均材料去除率,每个抛光过程对应唯一的材料去除率结

果。在剔除数据集中异常及不符合条件的晶片数据后,得到 2 551 条数据,其中训练集数据 1 777 条,测试集数据 383 条,验证集数据 391 条。

数据集中不同加工阶段和不同加工场所下晶片材料去除率的变化规律有明显不同,有必要针对不同的数据集场景分别建立模型。对数据集进行初步分析发现,以 MRR 等于 120 nm/min 为界,可以将数据分为 2 组:当 MRR 小于 120 nm/min 时,晶片均在场所 4、5、6 加工,且这些晶片有加工阶段 A、B 之分;当 MRR 大于 120 nm/min 时,晶片都在场所 1、2、3 加工,且都为阶段 A。

依据晶片加工阶段和加工场所可分为以下 3 组场景:1) CHAMBER_ID (加工舱室) 为 1、2、3, STAGE (加工阶段) 为 A;2) CHAMBER_ID (加工舱室) 为 4、5、6, STAGE (加工阶段) 为 A;3) CHAMBER_ID (加工舱室) 为 4、5、6, STAGE (加工阶段) 为 B。

1.2 特征提取

特征提取的目的是从连续过程信号中提取特征,以表征 MRR 的时变特征。对耗材使用量 (7-12)、压力 (13-18)、抛光液流量 (19-21)、转速 (21-24) 等参数提取基本统计特征,包括中位数 (Median)、标准差 (STD)、最小值 (Min)、最大值 (Max) 和平均值 (Mean) 等 5 个时域基本特征。此外,引入邻域特征以应对加工过程中动态因素的影响。根据 Jia 等^[10]的研究,本文将时间邻域特征、使用量邻域特征和压力邻域特征引入晶片平均材料去除率的预测中,以丰富原始数据特征、提升预测模型性能。

当前晶片的时间邻域特征与前几个晶片的材料去除率之间存在关联。在 3 种场景下,时间邻域特征与 MRR 的关系如图 1 所示。当前晶片的时间邻域特征提取过程如下:先按照加工顺序对所有晶片排序,

表 1 状态监测数据
Tab.1 Sensor data

Serial number	Signal	Serial number	Signal
1	MACHINE_ID	14	MAIN_OUTER_AIR_BAG_PRESSURE
2	MACHINE_DATA	15	CENTER_AIR_BAG_PRESSURE
3	TIMESTAMP	16	RETAINER_RING_PRESSURE
4	WAFER_ID	17	RIPPLE_AIR_BAG_PRESSURE
5	STAGE	18	EDGE_AIR_BAG_PRESSURE
6	CHAMBER	19	SLURRY_FLOW_LINE_A
7	USAGE_OF_BACKING_FILM	20	SLURRY_FLOW_LINE_B
8	USAGE_OF_DRESSER	21	SLURRY_FLOW_LINE_C
9	USAGE_OF_POLISHING_TABLE	22	WAFER_ROTATION
10	USAGE_OF_DRESSER_TABLE	23	STAGE_ROTATION
11	USAGE_OF_MEMBRANE	24	HEAD_ROTATION
12	USAGE_OF_PRESSURIZED_SHEET	25	DRESSING_WAFER_STATUS
13	PRESSURIZED_CHAMBER_PRESSURE		

丢弃初始 n 块晶片的数据; 在其后的晶片中添加 n 个特征分别对应之前 n 块晶片的平均材料去除率。

使用的量邻域特征与晶片加工过程中的耗材相关, 如抛光盘、背膜、平整器等。在 3 种场景下, 使用量邻域特征与 MRR 的关系如图 2 所示。使用量邻域特征的提取过程如下: 首先计算每个加工过程中晶片的平均耗材使用量, 在特征空间中, 若某样本附近的 K 个最近 (即特征空间中最邻近) 样本大多数属于一个类别, 则该样本也属于这个类别, 通过 K 近邻算法 (KNN, K -Nearest Neighbor, 一种简单的机器学习算法) 获取使用量最相近的晶片, 得到这些晶片的 MRR 后添加到当前晶片的特征值中, 作为使用量邻域特征。

压力邻域特征的提取过程如下: 首先计算得到每个晶片在加工过程中的平均压力和压力的标准差, 接着根据 K 近邻算法获取压力变化接近的晶片, 将得到的 K 个晶片的平均材料去除率添加到当前晶片的特征值中。3 种场景下当前晶片的 MRR 与压力邻域特征的相关性如图 3 所示。

考虑到晶片的平均材料去除率与抛光过程中压力作用在晶片上的时间相关, 将名为 “PRESSURE_KEEP_TIME” 的值引入晶片加工特征值中, 作为晶片和抛光盘之间实际接触时间的估算值。最后得到的数据以晶片 ID 为键, 分为压力、抛光液流速、使用量、转速、邻域特征和其他特征, 总共 95 个特征, 如表 2 所示。

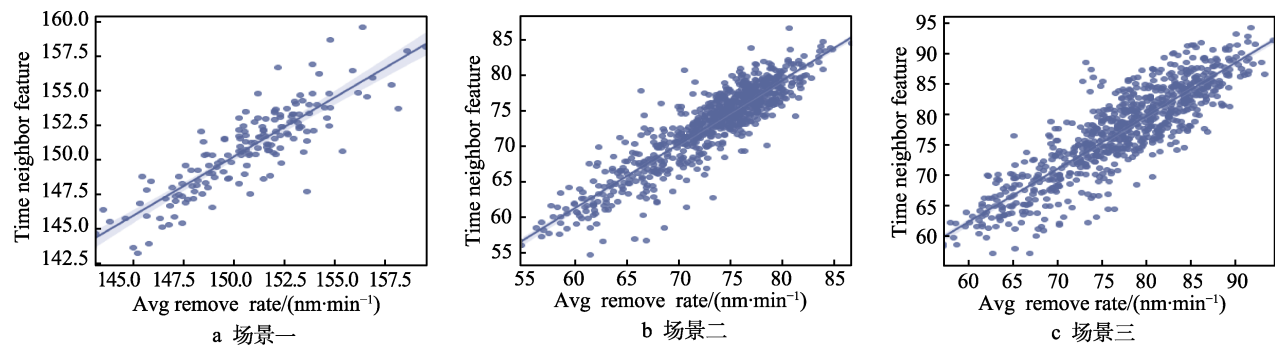


图 1 时间邻域特征与 MRR 相关图
Fig.1 Correlation plot of time neighborhood features and MRR:
a) scene 1; b) scene 2; c) scene 3

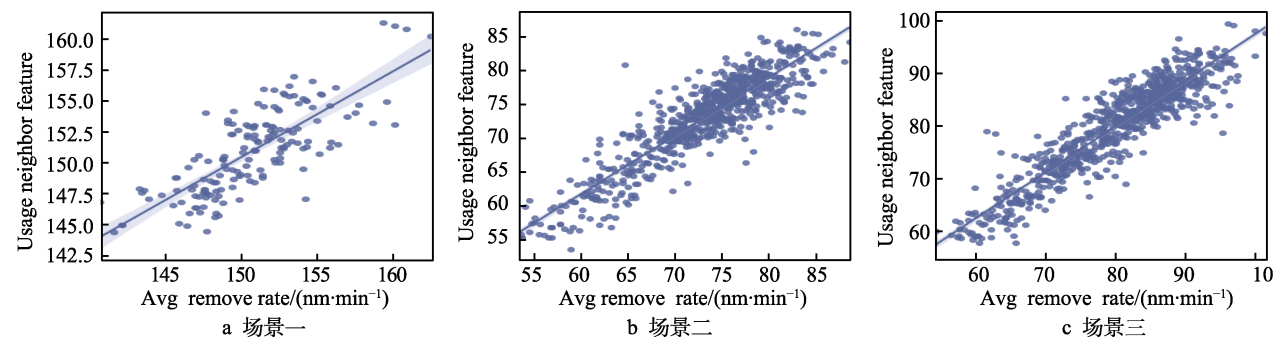


图 2 使用量邻域特征与 MRR 相关图
Fig.2 Correlation plot of usage neighborhood feature and MRR:
a) scene 1; b) scene 2; c) scene 3

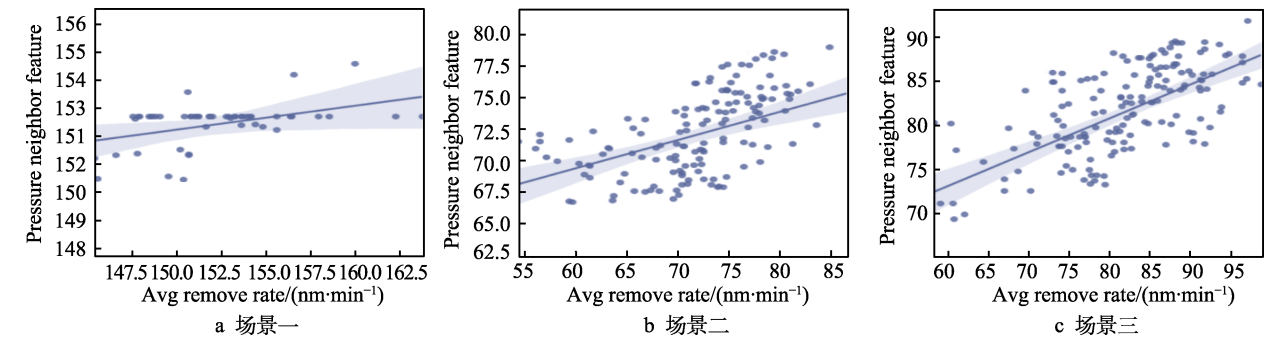


图 3 压力邻域特征与 MRR 相关图
Fig.3 Correlation plot of pressure neighborhood feature and MRR:
a) scene 1; b) scene 2; c) scene 3

表 2 提取的特征表
Tab.2 Extracted features

Grouping	Parameter name
Pressure	EDGE_AIR_BAG_PRESSURE (mean, std, min, max, median)
	RETAINER_RING_PRESSURE (mean, std, min, max, median)
	RIPPLE_AIR_BAG_PRESSURE (mean, std, min, max, median)
	CENTER_AIR_BAG_PRESSURE (mean, std, min, max, median)
	MAIN_OUTER_AIR_BAG_PRESSURE (mean, std, min, max, median)
	PRESSURIZED_CHAMBER_PRESSURE (mean, std, min, max, median)
Usage	USAGE_OF_DRESSER (mean, std, min, max, median)
	USAGE_OF_MEMBRANE (mean, std, min, max, median)
	USAGE_OF_BACKING_FILM (mean, std, min, max, median)
	USAGE_OF_DRESSER_TABLE (mean, std, min, max, median)
	USAGE_OF_POLISHING_TABLE (mean, std, min, max, median)
	USAGE_OF_PRESSURIZED_SHEET (mean, std, min, max, median)
Slurry flow rate	SLURRY_FLOW_LINE_A (mean, std, min, max, median)
	SLURRY_FLOW_LINE_B (mean, std, min, max, median)
	SLURRY_FLOW_LINE_C (mean, std, min, max, median)
Rotate speed	HEAD_ROTATION (mean, std, min, max, median)
	STAGE_ROTATION (mean, std, min, max, median)
	WAFER_ROTATION (mean, std, min, max, median)
Neighborhood feature	TIME_NEIGHBOR_FEATURE
	USAGE_NEIGHBOR_FEATURE
	PRESSURE_NEIGHBOR_FEATURE
Else	PRESSURE_KEEP_TIME
	DRESSING_WAFER_STATUS

2 基于 DAE 的 MRR 预测模型设计

2.1 模型搭建

基于 DAE 的 MRR 预测模型训练过程如图 4 所示, 首先进行数据的预处理及特征提取, 然后利用深度自动编码器对输入的特征数据进行融合, 然后将融合后的特征数据作为 BP 神经网络的输入, 训练整个 MRR 预测模型。

2.2 深度自动编码器模块设计

深度自动编码器 (Deep Autoencoder, DAE) 作为一种可以从数据中学习数据的无监督学习的人工神经网络, 常用于数据融合及数据降维^[19], 通过训练能过滤数据中的异常信息^[20], 以便后续回归模型的训练和预测。深度自动编码器还可用于主成分分析、图片处理^[21]、人气预测^[22]、机器翻译^[23]等方面, 分为编码器和解码器两部分^[24], 先输入数据 $\mathbf{x}=[x_1, x_2, \cdots, x_n]$, 待深度自动编码器运行后获取输出数据 $\mathbf{x}'=[x'_1, x'_2, \cdots, x'_n]$, 其中深度自动编码器隐藏层的输出为 $\mathbf{h}=[h_1, h_2, \cdots, h_d]$ 。DAE 编码器部分的公式如式 (1) 所示。

$$\mathbf{h} = f(\mathbf{W}\mathbf{x} + \mathbf{b}) \quad (1)$$

式中: \mathbf{x} 为输入数据; \mathbf{W} 为编码权重矩阵; \mathbf{b} 为编码偏置矩阵; f 为激活函数; \mathbf{h} 为隐藏层输出矩阵。

DAE 解码器部分的公式如式 (2) 所示。

$$\mathbf{x}' = f(\mathbf{W}'\mathbf{h} + \mathbf{b}') \quad (2)$$

式中: \mathbf{x}' 为解码输出; \mathbf{W}' 为解码权重矩阵; \mathbf{b}' 为

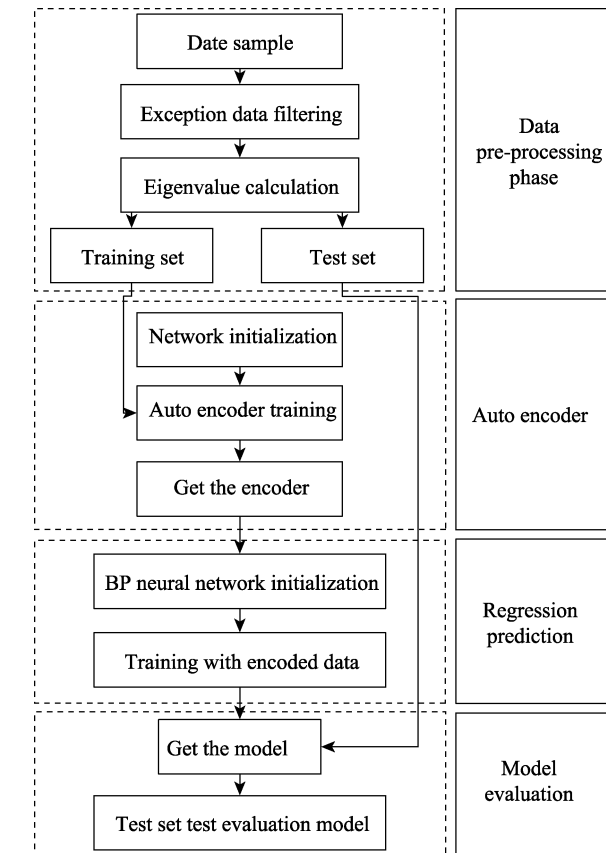


图 4 基于 DAE 的 MRR 预测模型
Fig.4 MRR prediction model based on DAE

解码偏置矩阵。

代价函数 J_{AE} 的表达式如式 (3) 所示。

$$J_{AE} = \frac{1}{2m} \sum_{i=1}^m \|x_i - x'_i\|_2^2 \quad (3)$$

式中: m 为训练的样本数。

深度自动编码器的训练和优化过程是通过最小化重建损失来执行的, 损失函数如式 (4) 所示。

$$L(x, x') = \|x - x'\|_2^2 = \|x - \sigma'(W'(\sigma(Wx + b)) + b')\|_2^2 \quad (4)$$

式中: σ 为隐藏层输出函数的函数符号; σ' 为输出层输出函数的函数符号。解码器中的 σ' 、 W' 、 b' 与编码器中的 σ 、 W 、 b 并无密切联系。

最小化损失函数用于提升神经网络模型性能, 使用动态加权损失函数迫使模型更加注重权重高的样本损失^[25]。为了提高强相关特征的重建精度, 解决不同特征值对晶片平均材料去除率预测的贡献不同的问题, 需要适当增加这部分特征损失函数数值。因此, 引入损失函数权重矩阵来调整不同特征之间的损失权重。引入自定义损失权重 W 后的代价函数如式 (5) 所示。

$$J_{(W)} = \frac{1}{N} \sum_{i=1}^N w_i (y_i - \hat{y}_i)^2 \quad (5)$$

式中: \hat{y} 为模型的预测值; y 为真实值; N 为样本数量。

损失权重根据特征值与 MRR 值的相关性确定为式 (6)。

$$w_i = 1 + \alpha \left(\rho_i - \frac{1}{N} \sum_{j=1}^N \rho_j \right) \quad (6)$$

式中: w_i 为单个特征值的权重; α 为权重系数; ρ 为特征值与 MRR 的相关性。

考虑到样本数据的完整程度, 在模型训练过程中引入与数据量和数据完整性有关的数据连续性系数 $C_{continuity}$, 以减少不完整数据的影响, 如式 (7) 所示。

$$C_{continuity} = \beta \lg \left(1 + \frac{t_{cur} - t_1}{\bar{t}} \frac{\sum_{i=2}^N (t_i - t_{i-1} - t_{AVG})}{t_N - t_1} \right) \quad (7)$$

式中: β 为权重系数; t_{cur} 为当前时间; t_{AVG} 为平均间隔时间; t_1 为加工开始时间; t_N 为加工完成时间; \bar{t} 为同场景下的平均加工时间。最终深度自动编码器的代价函数如式 (8) 所示。

$$J'_{(W)} = C_{continuity} J_{(W)} \quad (8)$$

2.3 BP 神经网络模块设计

BP 神经网络是最基本、应用最广泛的神经网络, 其训练方式为通过前向传播输出结果, 误差采用反向传播。

经过过滤去除一些数据后, 数据集仍然会存在一些异常数据。为减小这类数据对模型的影响, 确保模

型的整体性和预测准确性, 使用损失函数来防止模型拟合异常值, Huber Loss 是一种用于回归问题的带参损失函数, 如式 (9) 所示。

$$L_{\delta}(a) = \begin{cases} \frac{1}{2}a^2, & \text{for } |a| \leq \delta \\ \delta \cdot \left(|a| - \frac{1}{2}\delta \right), & \text{otherwise} \end{cases} \quad (9)$$

式中: a 为 y 与 $f(x)$ 的差值; $f(x)$ 为模型的预测值; δ 为常量。

当 δ 趋于 0 时, Huber 损失会趋于 MAE (Mean Absolute Error, 一种用于回归模型的损失函数) 损失; 当 δ 趋于正无穷时, Huber 损失趋于 MSE (Mean Square Error, 最常用的回归损失函数) 损失。与 MSE Loss 相比, Huber Loss 对异常数据的鲁棒性更大, 更稳健; 与最小二乘线性回归相比, Huber Loss 对离群点的惩罚程度更小。

HyperBand 是一种常见的超参数搜索算法, 它在 Jamieson 和 Talwalkar 提出的 Successive Halving 算法^[26]基础上做了扩展, 通过自适应资源分配和早停达到加速随机搜索的目的, 可以提前终止表现不好的超参数神经网络训练。本文运用 HyperBand 超参数搜索算法对 BP 神经网络的超参数进行了优化, 以解决超参数人为经验设定不可靠的问题。

3 基于 DAE 的 MRR 预测模型实验

3.1 消融实验

为验证基于 DAE 的 MRR 预测模型的效果, 以场景二数据集为例进行消融实验。首先使用未经 DAE 融合的特征数据训练 BP 神经网络基线模型。根据经验公式进行消融实验, 最终确定基线模型隐藏层神经元个数为 8, 并训练基线模型。模型训练过程中损失变化趋势如图 5 所示, 可以看到, 模型在 100 代后趋于稳定。鉴于模型使用了余弦退火学习率, 延长周期可能会找到更好的收敛点。为提升模型的预测精度和可靠性, 将模型的训练周期从 200 代提升到 1 000 代。由图 5 可知, 当模型从 600 代训练到 1 000 代时, 约有 1.5Loss 的下降, 虽然有继续降低的趋势, 但趋势不明显, 且损失数值已经足够小, 达到了较为优秀的水平, 可以停止训练。其次分别使用 Huber loss 和 Hyperband 超参搜索对基线模型进行优化, 实验结果如表 3 所示, 表中均方差误差 MSE 表示平均材料去除率的预测值与真实值之间的差异, 该值越小, 代表模型的预测效果越好。优化后的 MRR 预测模型在测试集上的 MSE 指标由 11.18 降低至 9.022, 预测精度明显提升。

为了验证深度自动编码器模型特征融合的性能, 根据 2.2 小节(深度自动编码器模块设计), 基于 DAE

的编解码结构, 首先将输入 $x \in R^d$ (d 维的实数向量) 编码到隐藏空间 $h \in R^d$ 中, 将高维数据编码到低维, 其次使用解码器在隐藏空间中将数据重建, 输出为 $x' \in R^d = x'$ 。解码器接收低维数据并尝试重建原始高维数据, 实现对输入特征的融合, 使特征具有更高的信息密度, 同时减少数据量。将 DAE 解码器输出的特征数据作为 BP 神经网络的输入数据, 进行 MRR 预测, 模型 MSE 降为 7.905, 见表 3。证明了基于 DAE 的 MRR 预测模型的有效性。

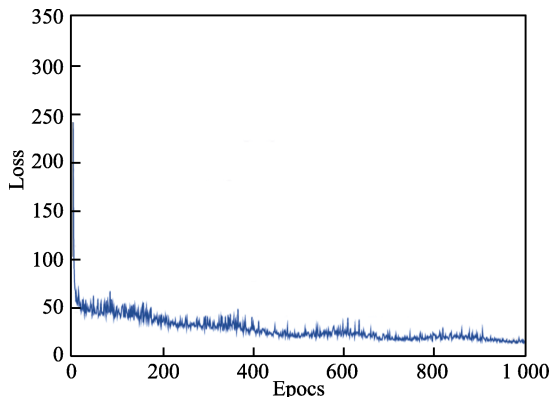


图 5 模型训练损失图
Fig.5 Model training loss graph

表 3 模型预测结果
Tab.3 Model prediction results

Model	MSE
Baseline	11.18
Baseline+Huber Loss ($\delta=12$)	10.63
Baseline+Huber Loss ($\delta=12$)+HyperBand algorithm	9.022
Baseline+Huber Loss ($\delta=12$)+HyperBand algorithm+DAE feature fusion	7.905

3.2 整体模型验证

线性回归模型、KNN 模型、BP 神经网络模型及本文模型在 3 组场景下数据集中训练的预测结果分别如图 6~8 所示。均方差误差 MSE 和平均绝对误差 MAE 表示平均材料去除率的预测值与真实值之间的差异, 其值越小, 代表模型的预测效果越好。相关系数 R^2 表示平均材料去除率的预测值与真实值之间的区别, 在 0~1 范围内, 该值越大, 代表模型的预测效果越好。

由图 6~8 知, 对于场景一的数据集, 虽样本较少, 但本文预测模型仍优于 BP 神经网络模型。对于场景二和场景三数据集, 本文模型在预测精度上远超过线性回归模型和 KNN 模型, 优于 BP 神经网络。通过对 BP 神经网络模型和本文模型进行对比观察, 发现在使用自编码器融合特征后, 本文模型的预测精度得到更高的提升, 在 3 个场景下的平均 MSE 损失达到 7.862, 相关性达到 91.2%。

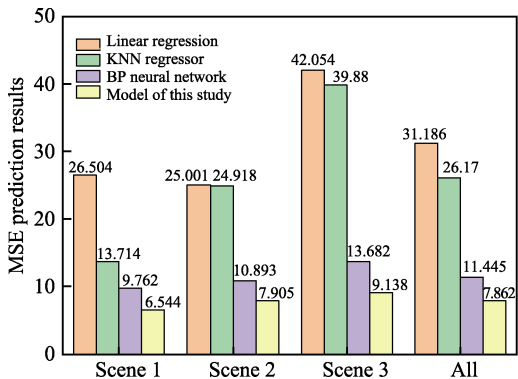


图 6 MSE 预测结果
Fig.6 MSE prediction results

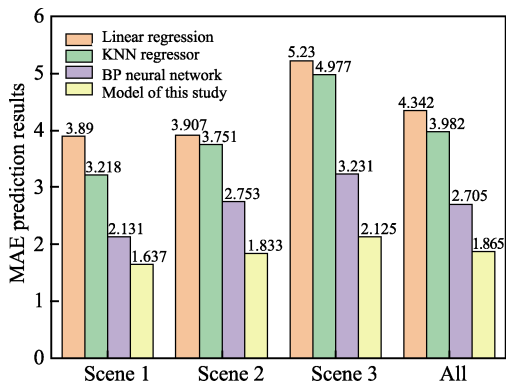


图 7 MAE 预测结果
Fig.7 MAE prediction results

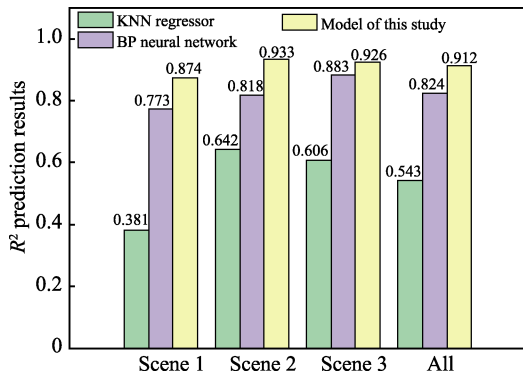


图 8 R^2 预测结果
Fig.8 R^2 prediction results

4 结论

利用晶片抛光过程中的压力、转速、耗材使用量、抛光液流速等数据, 建立了更加精确的材料去除率预测模型, 提出了基于多源数据融合的半导体基片抛光材料去除率预测方法。

1) 以 PHM2016 数据集为数据内容, 分析晶片抛光加工过程中的数据特点及数据之间的相关性。引入邻域特征来解决晶片加工过程中动态因素的影响。

2) 设计了深度自动编码器的参数, 优化了深度自动编码器的损失函数。在 BP 神经网络中, 使用

Hubor Loss 损失函数减小异常数据对模型的影响; 采用 HyperBand 超参数搜索算法对 BP 神经网络的超参数进行优化, 以解决超参数人为经验设定不可靠的问题。

3) 利用深度自动编码器融合多源传感器信号降低数据维度, 将融合后的数据输入 BP 神经网络中, 对晶片加工后的材料去除率进行预测, 提高了整体模型的 MRR 预测精度。

4) 相比于线性回归、KNN 模型和 BP 神经网络模型, 本文模型在 3 个场景下具有与实际值更接近的预测值, 能够更好满足半导体晶片抛光过程材料去除率的预测需求, 为抛光工艺的智能决策和控制奠定良好的基础。

参考文献:

- [1] SEO J. A Review on Chemical and Mechanical Phenomena at the Wafer Interface during Chemical Mechanical Planarization[J]. *Journal of Materials Research*, 2021, 36(1): 235-257.
- [2] CHEN H, GUO D, XIE G X, et al. Mechanical Model of Nanoparticles for Material Removal in Chemical Mechanical Polishing Process[J]. *Friction*, 2016, 4(2): 153-164.
- [3] 潘杰, 陈凡, 杨炜, 等. 基于 SPSO-BP 神经网络的自适应抛光工艺参数匹配[J]. *表面技术*, 2022, 51(8): 387-399.
- [4] PAN J, CHEN F, YANG W, et al. Adaptive Polishing Process Parameter Matching Based on SPSO-BP Neural Network[J]. *Surface Technology*, 2022, 51(8): 387-399.
- [4] XU Q Z, CHEN L, CAO H. A Wafer-Scale Material Removal Rate Model for Chemical Mechanical Planarization[J]. *ECS Journal of Solid State Science and Technology*, 2019, 8(12): 821-832.
- [5] 杜家熙, 苏建修, 万秀颖, 等. 单晶硅片化学机械抛光材料去除特性[J]. *北京科技大学学报*, 2009, 31(5): 608-611.
- [5] DU J X, SU J X, WAN X Y, et al. Material Removal Characteristic of Silicon Wafers in Chemical Mechanical Polishing[J]. *Chinese Journal of Engineering*, 2009, 31(5): 608-611.
- [6] 蒋建忠, 赵永武, 雒建斌. 一种基于非晶层粘性流动的机械化学抛光模型[J]. *中国机械工程*, 2006, 17(24): 2540-2546.
- [6] JIANG J Z, ZHAO Y W, LUO J B. A Chemical Mechanical Polishing Model Based on the Viscous Flow of the Amorphous Layer[J]. *China Mechanical Engineering*, 2006, 17(24): 2540-2546.
- [7] LIU D F, CHEN G L, HU Q. Material Removal Model of Chemical Mechanical Polishing for Fused Silica Using Soft Nanoparticles[J]. *The International Journal of Advanced Manufacturing Technology*, 2017, 88(9): 3515-3525.
- [8] YUH M, JANG S, KIM H, et al. Development of Green CMP by Slurry Reduction through Controlling Platen Coolant Temperature[J]. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 2015, 2(4): 339-344.
- [9] DI Y, JIA X D, LEE J. Enhanced Virtual Metrology on Chemical Mechanical Planarization Process Using an Integrated Model and Data-Driven Approach[J]. *International Journal of Prognostics and Health Management*, 2017, 8(2): 2153-2648.
- [10] JIA X D, DI Y, FENG J S, et al. Adaptive Virtual Metrology for Semiconductor Chemical Mechanical Planarization Process using GMDH-type Polynomial Neural Networks[J]. *Journal of Process Control*, 2018, 62: 44-54.
- [11] FENG J S, JIA X D, FENG Z, et al. An Online Virtual Metrology Model with Sample Selection for the Tracking of Dynamic Manufacturing Processes With Slow Drift[J]. *IEEE Transactions on Semiconductor Manufacturing*, 2019, 32(4): 574-582.
- [12] CAI H S, FENG J S, YANG Q B, et al. A Virtual Metrology Method with Prediction Uncertainty Based on Gaussian Process for Chemical Mechanical Planarization[J]. *Computers in Industry*, 2020, 119: 103228.
- [13] 游思奇. 解读机器学习与深度学习的发展及应用[J]. *计算机产品与流通*, 2018(5): 138.
- [13] YOU S Q. Explaining the Development and Application of Machine Learning and Deep Learning[J]. *Computer Products and Distribution*, 2018(5): 138.
- [14] 吴荣珍. 基于多源数据融合的传感器数据智能分析系统[J]. *太原师范学院学报*, 2022, 21(1): 51-56.
- [14] WU R Z. Intelligent Analysis System of Sensor Data Based on Multi-Source Data Fusion[J]. *Journal of Taiyuan Normal University*, 2022, 21(1): 51-56.
- [15] 郭创新, 彭明伟, 刘毅. 多数据源信息融合的电网故障诊断新方法[J]. *中国电机工程学报*, 2009, 29(31): 1-7.
- [15] GUO C X, PENG M W, LIU Y. Novel Approach for Fault Diagnosis of the Power Grid with Information Fusion of Multi-Data Resources[J]. *Proceedings of the CSEE*, 2009, 29(31): 1-7.
- [16] 姜延吉. 多传感器数据融合关键技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2010.
- [16] JIANG Y J. Research on Key Technologies of Multi-Sensor Data Fusion[D]. Harbin: Harbin Engineering University, 2010.
- [17] 胡昌华, 施权, 司小胜, 等. 数据驱动的使用寿命预测和健康管理技术研究进展[J]. *信息与控制*, 2017, 46(1): 72-82.
- [17] HU C H, SHI Q, SI X S, et al. Data-Driven Life Prediction and Health Management: State of the Art[J]. *Information and Control*, 2017, 46(1): 72-82.
- [18] JIA X D, HUANG B, FENG J S, et al. A Review of PHM Data Competitions from 2008 to 2017[C]// *Annual Conference of the PHM Society*, Philadelphia, US, 2018.

(下转第 167 页)

- Rotational Grinding[J]. Optics and Precision Engineering, 2022, 30(17): 2077-2087.
- [26] 段珊, 姜淑华, 王春阳. 光学元件亚表面损伤散射特性仿真研究[J]. 电光与控制, 2018, 25(9): 75-77.
DUAN S, JIANG S H, WANG C Y. Simulation Study on Sub-Surface Damage Scattering Characteristics of Optical Elements[J]. Electronics Optics & Control, 2018, 25(9): 75-77.
- [27] LI P H, GUO X G, YUAN S, et al. Effects of Grinding Speeds on the Subsurface Damage of Single Crystal Silicon Based on Molecular Dynamics Simulations[J]. Applied Surface Science, 2021, 554: 149668.
- [28] 王占奎, 朱永伟, 李信路, 等. 固结磨料研磨镁铝尖晶石的平均切深和亚表面损伤行为[J]. 硅酸盐学报, 2017, 45(3): 402-409.
WANG Z K, ZHU Y W, LI X L, et al. Average Cutting Depth and Subsurface Damage of Spinel Induced by Lapping with Fixed Abrasive Pad[J]. Journal of the Chinese Ceramic Society, 2017, 45(3): 402-409.
- [29] 戴子华, 朱永伟, 王建彬, 等. K9 玻璃亚表面损伤的分步腐蚀法测量[J]. 光学精密工程, 2013, 21(2): 287-293.
DAI Z H, ZHU Y W, WANG J B, et al. Measurement of Sub-Surface Damage of K9 Glass by Step-by-Step Etching Method[J]. Optics and Precision Engineering, 2013, 21(2): 287-293.
-
- (上接第 149 页)
- [18] WANG J D, CHENG C, ZENG X T, et al. Bionic-Tribology Design of Tooth Surface of Grinding Head Based on the Bovine Molar[J]. Tribology International, 2019, 143: 106066.
- [19] GUO Y C, LEE H, LEE Y, et al. Effect of Pad Groove Geometry on Material Removal Characteristics in Chemical Mechanical Polishing[J]. International Journal of Precision Engineering and Manufacturing, 2012, 13(2): 303-306.
- [20] 王君明, 叶人珍, 汤漾平, 等. 55 钢 CBN 砂轮平面磨削的磨削力模型研究[J]. 金刚石与磨料磨具工程, 2010, 30(1): 67-70.
WANG J M, YE R Z, TANG Y P, et al. Research on the Grinding Force Model of Steel 55 during Surface Grinding with CBN Wheel[J]. Diamond & Abrasives Engineering, 2010, 30(1): 67-70.
- [21] AURICH J C, KIRSCH B, HERZENSTIEL P, et al. Hydraulic Design of a Grinding Wheel with an Internal Cooling Lubricant Supply[J]. Production Engineering, 2011, 5(2): 119-126.
- [22] 王文珊. 蓝宝石加工用图案化凝胶研磨盘的设计制备及加工机理研究[D]. 泉州: 华侨大学, 2020.
WANG W S. Design, Preparation and Processing Mechanism of Patterned Gel Grinding Disk for Sapphire Processing[D]. Quanzhou: Huaqiao University, 2020.
- [23] 石兴泰, 郭磊, 刘晓辉, 等. 随机网格结构固结磨料磨盘平面磨削性能研究[J]. 金刚石与磨料磨具工程, 2022, 42(3): 275-282.
SHI X T, GUO L, LIU X H, et al. Study on Machining Performance of Fixed-Abrasive Lap-Grinding Plate with Random Grid Structure[J]. Diamond & Abrasives Engineering, 2022, 42(3): 275-282.
-
- (上接第 157 页)
- [19] CHICCO D, SADOWSKI P, BALDI P. Deep Autoencoder Neural Networks for Gene Ontology Annotation Predictions[C]// ACM BCB 2014-5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, 2014: 533-540.
- [20] MORALES-FORERO A, BASSETTO S. Case Study: A Semi-Supervised Methodology for Anomaly Detection and Diagnosis[C]// 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE, 2019: 1031-1037.
- [21] MUKHERJEE R, CHANDRAN S. Lossy Image Compression Using SVD Coding, Compressive Autoencoders, and Prediction Error-vector Quantization[C]// 2017 4th International Conference on Opto-Electronics and Applied Optics (Optronix), IEEE, 2017: 1-5.
- [22] DE S, MAITY A, GOEL V, et al. Predicting the Popularity of Instagram Posts for a Lifestyle Magazine using Deep Learning[C]// 2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA), IEEE, 2017: 174-177.
- [23] CHO K, MERRIENBOER B, BAHDANAU D, et al. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches[J]. Computer Science, 2014.
- [24] 刘会芸, 侯志平. 基于自动编码器数据降维的滚动轴承故障诊断研究[J]. 现代制造工程, 2022(1): 148-153.
LIU H Y, HOU Z P. Research on Rolling Bearing Fault Diagnosis Based on Auto-Encoder Data Dimensionality Reduction[J]. Modern Manufacturing Engineering, 2022 (1): 148-153.
- [25] RENGASAMY D, JAFARI M, ROTHWELL B, et al. Deep Learning with Dynamically Weighted Loss Function for Sensor-Based Prognostics and Health Management[J]. Sensors, 2020, 20(3): 723.
- [26] JEONG D, LEE S, LEE J H. Improving Successive Halving Algorithm by Considering the Learning Characteristics of Neural Network[J]. Journal of Korean Institute of Intelligent Systems, 2019, 29(5): 333-338.