基于堆叠降噪自编码器的神经–符号模型及在晶圆表面缺陷识别

刘国梁 [1] 余建波 [1]

摘 要 深度神经网络是具有复杂结构和多个非线性处理单元的模型, 通过模块化的方式分层从数据提取代表性特征, 已经在晶圆缺陷识别领域得到了较为广泛的应用. 但是, 深度神经网络在应用过程中本身存在 "黑箱"和过度依赖数据的问题, 显著地影响深度神经网络在晶圆缺陷识别的工业可应用性. 提出一种基于堆叠降噪自编码器的神经–符号模型. 首先, 根据堆叠降噪自编码器的网络特点采用了一套符号规则系统, 规则形式和组成结构使其可与深度神经网络有效融合. 其次, 根据 网络和符号规则之间的关联性提出完整的知识抽取与插入算法, 实现了深度网络和规则之间的知识转换. 在实际工业晶圆表面图像数据集 WM-811K 上的试验结果表明, 基于堆叠降噪自编码器的神经–符号模型不仅取得了较好的缺陷探测与识别性能, 而且可有效提取规则并通过规则有效描述深度神经网络内部计算逻辑, 综合性能优于目前经典的深度神经网络.

关键词 晶圆表面缺陷, 深度学习, 堆叠降噪自编码器, 符号规则, 知识发现

引用格式 刘国梁, 余建波. 基于堆叠降噪自编码器的神经–符号模型及在晶圆表面缺陷识别. 自动化学报, 2022, **48**(11): 2688−2702

DOI 10.16383/j.aas.c190857

Application of Neural-symbol Model Based on Stacked Denoising Auto-encoders in Wafer Map Defect Recognition

LIU Guo-Liang[1] YU Jian-Bo[1]

Abstract Deep neural network is a model with complex structure and multiple non-linear processing units. It has achieved great successes in wafer map pattern recognition through deep feature learning. In order to solve the problem of unexplained "black box" and excessive dependence on data in the applications of deep neural networks, this paper proposes a neural-symbol model based on a stacked denoising auto-encoders. Firstly, the symbolic rule system is designed according to the characteristics of stacked denoising auto-encoders. Secondly, according to the inner association between the network and the rules, a knowledge extraction and insertion algorithm is proposed to describe the deep network and improve the performance of the network. The experimental results on the industrial wafer map image set WM-811K show that the neural-symbol model based on stacked denoising auto-encoders not only achieves better defect pattern recognition performance, but also can effectively describe the internal logic of the neural network through rules, and its comprehensive performance is better than that of the current classical classification model.

Key words Wafer map defect, deep learning, stacked denoising auto-encoders, symbolic rule, knowledge discovery

Citation Liu Guo-Liang, Yu Jian-Bo. Application of neural-symbol model based on stacked denoising auto-encoders in wafer map defect recognition. *Acta Automatica Sinica*, 2022, **48**(11): 2688−2702

半导体作为应用最为广泛的元器件之一, 其制造过程需要经过薄膜沉积、蚀刻、抛光等众多复杂工艺流程, 生产过程中的任何异常都可能导致晶圆表面缺陷的产生[1]. 除了需要对晶圆制造过程中的关键参数进行控制和预测[2], 准确识别晶圆表面的各种缺陷模式, 也有助于提升晶圆制造质量, 降低半导体生产废品率, 避免因大批量晶圆表面缺陷而造成的巨大损失.

早期的晶圆表面缺陷识别方法主要通过统计学方法实现. Hess 等[3] 研究晶圆缺陷密度分布实现对成品率的预测. Friedman 等[4] 采用无模型的缺陷聚类方法实现对晶圆表面缺陷的形状、大小和分布的检测. Yuan 等[5] 在前人研究的基础上提出一种基于贝叶斯推论的模式聚类演算法, 可进一步检测曲

收稿日期 2019-12-17 录用日期 2020-05-18
Manuscript received December 17, 2019; accepted May 18, 2020
国家自然科学基金 (71771173) 资助
Supported by National Natural Science Foundation of China (71771173)
本文责任编委 胡清华
Recommended by Associate Editor HU Qing-Hua
1. 同济大学机械与能源工程学院 上海 201804
1. School of Mechanical and Energy Engineering, Tongji University, Shanghai 201804

线模式、椭球模式、非均匀全局缺陷模式. 这些方法的缺陷在于只是对晶圆缺陷进行了统计分析, 并没有做到对缺陷类别的精准识别, 对实际生产过程帮助有限.

随着机器学习和深度学习的崛起, 线性判别方法[6]、反向传播网络[7]、广义回归神经网络[8]、支持向量机[8−10]、深度神经网络[11−14] 等模型已被广泛地应用于晶圆表面缺陷识别, 其中堆叠降噪自编码器 (Stacked denoising auto-encoders, SDAE) 作为经典的深度学习模型, 凭借其强大的学习能力, 取得了不错的结果[13−14]. 但是, 上述模型仍然存在以下 2 个问题: 1) 虽然以卷积神经网络和 SDAE 为代表的深度神经网络模型凭借其强大的特征提取能力, 在晶圆缺陷识别问题上取得了较好的结果, 但是深度网络模型始终存在不可被解释的缺陷. 这一缺陷使得深度神 经网络在 WMPR 上的应用存在很多困难. 2) 传统机器学习模型如支持向量机、决策树等可以通过数学或逻辑途径进行解释和验证, 但是它们的缺陷识别能力并不高.

纵观神经网络发展史, 研究者们一直在尝试弥补神经网络不可被解释的缺陷. 通过对网络的结构参数或统计意义进行分析, 以达到解释网络的目的是当下的主流研究方向[15]. Gallant[16] 最早提出利用 IF-THEN 形式的规则解释神经网络的推理结果, 形成神经网络专家系统. 其后 Towell 等[17] 提出基于知识的人工神经网络 (Knowledge-based artificial neural network, KBANN), 该模型通过从网络中抽取和插入规则, 实现了逻辑规则与神经网络之间的交互. Garcez 等[18] 在 KBANN 的研究基础上提出一种利用符号规则初始化神经网络的方法, 可以帮助模型更高效的学习数据中的知识. 在深度神经网络研究方面, Garcez 等[19] 提出神经–符号系统的概念, 其核心理念为符号规则负责表述神经网络中蕴含的知识而神经元负责学习和推理, 所生成的模型同时具备高鲁棒性、高识别性能以及可解释性. 在这一概念的基础上, Odence 等[20] 将受限玻尔兹曼机与符号规则相结合, 为符号规则与深度神 经网络的结合打下基础; Tran 等[21] 在前人研究基础上首次提出了从深度置信网络 (Deep belief network, DBN) 中抽取和插入符号规则的算法, 具有里程碑意义; 刘国梁等[22] 提出一种混合规则并将它与堆叠降噪自编码器集成, 但该算法计算成本高, 难以适应大规模复杂问题, Hitzler 等[23] 在符号–神经系统的基础上, 详细介绍语义网的神经符号研究的前景和优势, 并分析了其对深度学习的潜在场景. Bennetot 等[24] 提出了一种推理模型来解释神经网络的决策, 并使用解释从网络原理来纠正其决策过程种的偏差. 在推理模型方面: Li 等[25] 从功能角度将逻辑语言与神经网络相结合, 形成了一种新的学习推理模型, 同时具备连接主义和符号主义的优势. Sukhbaatar 等[26] 提出了记忆网络, 引入了记忆机制来解决对推理过程中间结果的存储问题, 对神经符号系统进行了进一步的探索, 赋予了神经网络符号化的结构, 对后续的研究有着重要的启发意义. Sawant 等[27] 在知识图和语料库的基础上建立了一套推理系统, 可以解释模型中不可观察或潜在的变量. Liang 等[28] 进一步引入了符号化的记忆机制, 帮助神经网络更好地完成复杂推理. Salha 等[29] 利用简单的线性模型替代图自编码器等模型中的图卷积网络, 简化了模型计算. 同时, Salha 等[30] 提出了一个通用的图自编码器和图变分自编码器的框架. 该框架利用图的简并性概念, 只从密集的节点子集中训练模型, 从而显著提高了模型的可伸缩性和训练速度. 综上所述, 目前对传统深度学习模型 (比如 DBN 或 SDAE) 的可解释性研究已经初见成效, 但在卷积神经网络类网络中, 卷积等运算带来的复杂问题在可解释性上还有待研究. 如何建立一套适用于晶圆缺陷识别的神经–符号模型是本文研究的重点.

针对晶圆缺陷识别问题的特点, 基于神经与符号相结合的理念, 本文采用一种基于 SDAE 的神经–符号模型[22], 构建了基于知识的堆叠降噪自编码器 (Knowledge-based stacked denoising auto-encoder, KBSDAE), 并建立了一套基于 KBSDAE 的晶圆表面缺陷识别系统, 以达到快速、高效识别晶圆表面缺陷的目的. 本文的主要贡献包括: 1) 提出了全新的符号规则形式, 可有效地表达 SDAE 的深度网络结构, 极大地减少了知识转化过程中的信息损失; 2) 提出了规则抽取与插入算法, 在实现知识高效转化的同时提升 SDAE 特征学习性能; 3) 提出了基于神经–符号系统的晶圆缺陷识别模型, 既可以识别缺陷模式, 也可以通过规则理解网络内部的推理逻辑, 并使得深度神 经网络具有了可解释性. 基于 SDAE 的神经–符号系统成功应用在实际工业案例中且取得了较好的特征学习和识别性能, 是在晶圆表面缺陷识别领域的一次新的尝试.

1　堆叠降噪自编码器

自编码器由输入层 (x)、隐藏层 (h) 和输出层 (y) 构成, 是深度学习的经典模型之一[1]. 它通过编码和解码运算重构输入数据, 通过减少重构误差为目标达到特征提取的目的. 由于训练过程中没有利用数据标签, 而只是以输入数据作为重构目标, 属于典型的无监督学习.

自编码器的编码阶段在输入层 x 和隐藏层 h 之

间, 具体表示为:

$$h = f_\theta(x) = \sigma(wx + b) \tag{1}$$

式中, σ 是非线性激活函数 Sigmoid 函数: $\sigma(x) = 1/(1 + \mathrm{e}^{-x})$, 参数集合 $\theta = \{w, b\}$. 解码阶段体现在隐藏层 h 和输出层 y 之间, 表示为:

$$y = g_{\theta'}(h) = \sigma'(w'h + b') \tag{2}$$

式中, σ' 是非线性激活函数 Sigmoid 函数, 参数集合 $\theta' = \{w', b'\}$.

通过最小化重构误差函数 $L(x, y) = \|x - y\|^2$ 来逐步地调整网络内部的参数 θ 和 θ', 优化方式选择随机梯度下降法, 最优参数如下:

$$(\theta, \theta') = \arg\min_{\theta, \theta'} L(x, g_{\theta'}(f_\theta(x))) \tag{3}$$

降噪自编码器 (Denoising auto-encoder, DAE) 是基于自编码器的一种变形, 通过噪声污染训练输入数据以增加网络的鲁棒性, 防止过拟合[31]. 图 1 展示了 DAE 的训练过程, 首先利用随机函数以一定的概率 p 将原训练数据 x 中的一些单元置零得到被污染的数据 $\tilde{x}$; 其次通过自编码器对 $\tilde{x}$ 进行重构; 最后调整网络参数 θ 和 θ'. DAE 相较于传统的自编码器具有更强的泛化能力和鲁棒性.

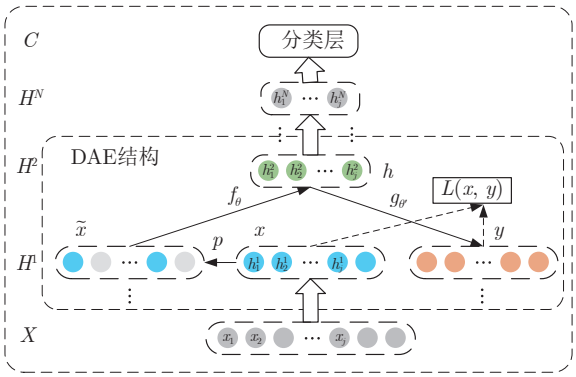


图 1 堆叠降噪自编码器

Fig. 1 Stacked denoising autoencoder

将若干个 DAE 堆叠起来, 就可以形成 SDAE, 如图 1 所示. 其训练过程首先是对逐个 DAE 进行训练, 其次通过反向传播算法微调整个网络. 相较于浅层神经网络, 层度更深的 SDAE 在特征提取方面更加优秀, 在处理高维数据问题上具有明显优势. 从符号与网络相结合的角度来看, 它的网络结构简单并且支持将 Sigmoid 作为激活函数, 这两个特性使 SDAE 更容易与符号规则进行集成.

2 神经–符号规则系统

符号规则的应用不仅能够实现对网络的描述和解释, 还能够提高模型性能. 本节主要讨论 SDAE 与符号规则结合建立模型的方法. 如图 2 所示, 该模型的建立分为 3 步: 1) 建立并训练标准 SDAE; 2) 从 SDAE 中抽取知识得到符号规则与分类规则; 3) 将符号与分类规则插入 SDAE 进行深度学习. 符号规则和神经网络的集成可实现二者优势的互补, 规则可以描述网络并表达深度网络中的知识, 而 KBSDAE 可以更有效地识别晶圆缺陷.

2.1 符号规则系统

以往逻辑符号规则种类繁多, 但都有同样的缺点, 即表现形式和推理逻辑单一. 这一缺点导致传统规则在描述参数庞大的深度网络时会出现规则体积庞大、描述效率底下和难以推导并理解的问题. 针对 SDAE 的网络特点, 本文在传统规则的基础上提出了一种数值和符号相结合的规则系统, 解决 SDAE 不能被解释的问题.

作为一种符号语言, 规则的形式对规则本身意义重大, 合适的形式才能更高效表示和描述网络. 由于 SDAE 包含特征提取部分的降噪自编码器 (Denoising auto-encoders, DAEs) 和用于分类的分类器, 虽然 2 部分的形式相同, 但是运行机理截然不同. 为了能更精准地描述网络, 根据网络不同部分的特性确定了不同的规则形式: 置信度规则和 MofN (N 个先行条件中的 M 个为真) 规则, 并将它们有机地结合起来.

网络特征提取部分由多个 DAE 叠加形成, 其训练方式为逐层训练. 为了保证置信度规则能够有效描述网络的这一部分, 置信度规则具备了以下特性[21−22]: 规则本身支持逐层推导; 规则节点与网络神经元一一对应; 置信值是对网络权值进行拟合得出的; 推理过程由符号和数值共同完成.这些特性赋予符号规则 3 种能力: 1) 规则具备描述大型网络的能力, 且逐层推导的逻辑意义与 DAEs 部分一致; 2) 符号规则的结构与网络基本相同且元素一一对应, 网络内部的逻辑关系可以被迁移到规则上作为一种网络内部关系的表现; 3) 规则可以作为深度神经网络的一种简化表示, 具备一定的识别能力. 所以符号规则的运行其实是对神经网络行为的一种简化模仿, 而这种模仿过程是人类所能理解的.

置信度规则[21] 是一个符合充要条件的等式: $c: h \leftrightarrow x_1 \wedge \cdots \wedge x_n$, 其中 c 是实数类型, 定义为置信值; h 和 $x_i (i \in [1, n])$ 为假设命题. 这种符号规则形式与文献 [21] 的规则相似, 但由于面向的网络不同, 规则符号的意义也不同. 本文定义具体的置信度符号规则:

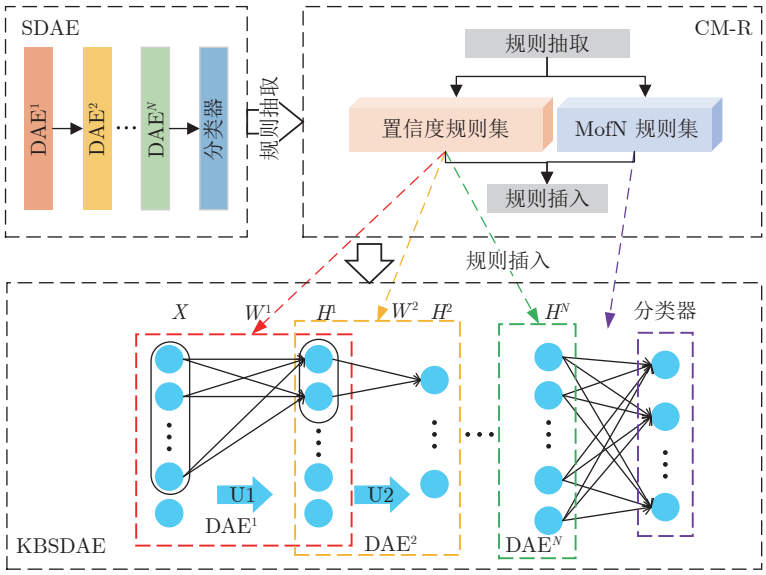


图 2　堆叠降噪自编码器的神经–符号模型

Fig. 2　Stacked denoising autoencoder based neural-symbolic model

$$\lambda^l=\begin{cases}c_j^l:h_j^l\leftrightarrow(\wedge_{\forall p\in P}h_p^{l-1})\wedge(\wedge_{\forall n\in N}\neg h_n^{l-1}),\\ \qquad 1<l<N\\ c_j^l:h_j^l\leftrightarrow(\wedge_{\forall p\in P}x_p)\wedge(\wedge_{\forall n\in N}\neg x_n),\\ \qquad l=1\end{cases}\tag{4}$$

该规则被解释为: 当 $x_1,\cdots,x_n$ 命题成立时, h 命题也成立的置信值为 c, 反之也成立. 其中 λ_j^l 是符号规则标签, 解释为第 l 层第 j 个符号规则; h_j^l 代表 DAE 中第 l 个隐藏层中第 j 个神经元; $x_i(i\in[1,n])$ 代表 DAE 输入层中第 i 个神经元, P 和 N 分别代表对 h_j^l 产生积极和消极影响的输入层神经元集合. 根据表达式可以看出, 置信度规则和 DAEs 具有相似的堆叠嵌套结构, 这可以最大化模拟网络结构.

SDAE 的分类器层一般为单层前向全连接网络, 通过反向传播算法进行训练. 这种经典网络的规则模型研究较为成熟, 故本文采用 Towell 等[32]提出的 MofN 规则形式. 这种规则通过对网络权重值和偏差的归纳与总结, 达到从网络中抽取规则的目的. 相较于同类型的其他规则, MofN 具备形式灵活和体积小的优点, 这使得它可以适用于规模较大的网络. 分类规则的基本表达形式如下:

IF (如果 N 个先行项中的 M 个为真) THEN $\cdots$　(5)

该规则被解释为: 如果规则的 N 个前层神经元中有 M 个被激活, 那么这条规则所对应的神经元也激活. 为了使 MofN 规则与置信度规则更加契合, 使用式 (5) 的泛化形式:

$$\text{IF }(w_1\times NumTrue(A_1)+\cdots+ w_n\times NumTrue(A_n)>bias)\text{ THEN }y=C\tag{6}$$

式中, $NumTrue$ 代表神经元激活的数量; A 代表一类前层神经元的集合, w 代表一类连接的权重值, 类别通过对权重值聚类得到; $bias$ 代表目标神经元的偏置值; C 表示具体的类标签.

上述 2 种规则的有机结合形成了一套规则体系 (Confidence & MofN rule, CM-R)[22], 具备以下优点: 规则本身具备分层特性, 可进行逐层抽取和推导, 与 SDAE 的堆叠逻辑相通; 规则根据网络不同部分的不同特性有针对的进行设计, 极大地减少了抽取过程中的信息损失; 这两种规则的集成使 CM-R 在处理复杂数据时也具有较高准确度.

CM-R 可逐层推理的特性是其能够适配 SDAE 的根本原因, 也是置信度规则和 MofN 规则可以集成的根本因素, 所以规则层与层之间的推理方法是极为重要的. 本文根据将规则的数值特性和符号特性相结合, 提出了一套适用于 CM-R 的推理算法 (Rule inference, Rule-INF)[22]. Rule-INF 以符号结构作为规则层内推导依据, 以数值作为层与层之间的联系, 将整个 CM-R 联系了起来, 使之成为一个完整的规则系统. 这一算法最大特点是通过对置信值的推导使规则突破了离散二值的限制, 可以被用来推导连续数据. 算法细节如下所示, 首先将初始化后的数据输入置信度规则中进行逐层推导, 其中上层规则推导输出的信任值 (B) 可作为下层规则的输入数据; 其次将顶层置信度规则输出的信任值调整为布尔向量; 最终利用 MofN 规则根据调整后的信任值 (1 表示真、0 表示假) 确定数据类别.

算法 1. Rule-INF

输入. CM-R rule set, dataset X.

输出. The result of inference.

1) $X \leftarrow Norm(X)$ // 归一化;

2) For $l = 1$ to N do // N 是隐藏神经元数目;

3) $B^l = \{\}$ // Initialize confidence vector;

4) For each symbolic rule λ_j^l do;

5) $\alpha_t \leftarrow X_t$, $\alpha_k \leftarrow X_k$ ($t \in T$, $k \in K$);

6) $\alpha = c_j^l \cdot (\sum_t \alpha_t - \sum_k \alpha_k)$;

7) $B_j^l \leftarrow Norm(\alpha)$;

8) End for;

9) $X \leftarrow B^l$;

10) End for;

11) For each element in B^N do // $B^N = B^{l=N}$;

12) IF $B_i^N > Random(1)$ THEN $B_i^N = 1$ else $B_i^N = 0$ // $Random(1)$ generate random number in 0 ~ 1;

13) End for;

14) For each rule do;

15) IF $NumTrue(B^N) \cdot w > bias$ THEN $y = c$; // Infer Mof-N according to B^N and get the final result;

16) End for.

2.2 知识抽取

本节将呈现从 SDAE 模型中抽取规则. 由于符号规则 CM-R 是知识的载体, 故知识抽取也叫规则抽取. CM-R 包含置信度规则和 MofN规则, 分别对应 SDAE 中的 DAE 和分类器部分, 下面对 2 种规则进行讨论.

置信度规则面向特征提取部分[21] 有逐层无监督训练和多个 DAE 堆叠而成 2 个特点. 为了使知识抽取过程更加符合网络的训练逻辑, 引入了逐层抽取的概念, 即在自监督训练过程中对每一个 DAE 单独抽取规则. 规则抽取原理是将置信值 $c_j s_j$ 最大化拟合权重值 w_j, 并利用符号解释网络结构. 根据 DAE 基本原理, 其输入数据 x 到隐含表示 h 的映射表示为:

$$h_j = \sigma(w_j^{\mathrm{T}} x - b_j) \tag{7}$$

式中, σ 表示激活函数 Sigmoid, b 表示偏置值. 根据式 (7), 本文提出新的函数, 可将数据 x 映射到隐藏层空间中:

$$h'_j = \sigma(c_j s_j^{\mathrm{T}} x - b_j) \tag{8}$$

式中, c_j 是连续实数, $s_j \in \{1, 0, -1\}$ 是对 w_j 的符号项表示, 可以理解为 $s_{ij} = \mathrm{sgn}(w_{ij})$. 对比式 (7) 和式 (8) 可以看出, 2 个公式的形式和元素基本相同. 为了让 h'_j 可以有效地代表隐藏层空间, 需要找到合适的 c_j 和 s_j 使 h'_j 近似于 h_j. 本文通过最小化 w_j 与 $c_j s_j$ 之间的欧氏距离实现拟合过程:

$$d(w, cs) = \sum_{ij} \|w_{ij} - c_j s_{ij}\|^2 \tag{9}$$

经过数学推导, 最终可以得到 c_j 的表达式:

$$c_j = \frac{\sum\limits_i w_{ij} s_{ij}}{\sum\limits_i s_{ij}^2} \tag{10}$$

进一步分析式 (9) 可知:

$$\|w_{ij} - c_j s_{ij}\|^2 = \begin{cases} (|w_{ij}| + c_j)^2, & s_{ij} = -1 \\ (|w_{ij}| - c_j)^2, & s_{ij} = 1 \\ |w_{ij}|^2, & s_{ij} = 0 \end{cases} \tag{11}$$

在 $s_{ij} = 1$ 或 $s_{ij} = 0$ 的情况下式 (9) 最小, 当且仅当 $|w_{ij}|^2 < (|w_{ij}| - c_j)^2$ 时 $s_{ij} = 0$ 条件会使信息损失最小. 所以对一个规则 j, 如果输入 x_i 对应的权重符合当 $2 \times |w_{ij}| \le c_j$ 条件, 那么该输入需要被删除. 通过上述筛选算法, 可以得到一种具有强连接关系的置信值的规则[21], 可有效地描述 DAEs.

算法 2. 置信度符号规则抽取.

输入. 1 个训练好的 DAE 模型.

输出. 信任规则集.

1) For $j = 1$ to the number of hidden units do;

2) Create sign matrix s with each $s_{ij} = \mathrm{sign}(w_{ij})$;

3) $c_j = \sum_{s_{ij}=0} |w_{ij}| / \sum_i s_{ij}^2$;

4) For each $s_{ij} \neq 0$ do;

5) IF $2 \times |w_{ij}| \ge c_j$ THEN;

6) Add x_i or $\neg x_i$ into rule r_j;

7) End for;

8) Until the value c_j is unchanged;

9) End for.

根据上述分析, 从 DAEs 中抽取置信度符号规则的置信度符号规则抽取 (Confidence rule extraction, Confidence-RE) 如算法 2 所示. 该算法面向单个 DAE, 所以只需根据网络将其迭代运行, 抽取出完整且具有堆叠特性的置信度规则集[22].

MofN 规则[25] 面向 SDAE 的分类器部分, 本文仅讨论以单层神经元为分类器的网络, 后文用分类层表述这一单层神经网络. 在进行规则抽取之前首先要对网络的微调过程进行假设: 分类层和隐藏层 H^N (如图 1 所示) 只具备激活 (输出值接近 1) 和不激活 (输出值接近 0) 两种输出状态. 这一假设使得分类器相关的神经元具备布尔特性, 把规则抽取问题转化成了神经元是否激活的规律性问题.

为了符合上述假设, 将逻辑回归函数作为激活

函数对网络进行微调. 分类层的微调原理为:

$$C_j = \sigma\left(\sum_i (w_{ji} \times x_i) - b_j\right) \tag{12}$$

式中, C_j 表示分类层中第 j 个神经元, 逻辑回归函数 σ 表示为:

$$\sigma(x) = \frac{1}{1+\mathrm{e}^{-x}} \tag{13}$$

由式 (12) 和式 (13) 可知, 当神经元的加权输入值大于偏置值时, 其输出值接近 1, 反之则接近 0. 这与假设相匹配. MofN 的规则抽取过程可以看作是搜索使分类层神经元激活的条件情况.

分类器部分神经元的输出值被简化成了 0 或 1, 使得神经元的输入被简化成只与权重值有关, 式 (12) 可简化为:

$$C_j = \sigma\left(\sum_{i|x_i=1} w_{ji} - b_j\right) \tag{14}$$

这一简化使规则抽取只需关注分类层神经元的连入权重和自身的偏置, 显著降低规则和算法复杂度.

MofN 规则抽取算法分为 4 步: 1) 通过 K 均值将分类层神经元的连入权重值聚类并将组内成员的权重值重置为组标签; 2) 对神经元影响不大的权重类删除 (归零); 3) 固定权重值, 通过反向传播算法重新对神经元偏置进行优化; 4) 对每一个分类层神经元形成一条规则, 其中神经元偏置作为阈值, 权值连接的 H^N 层神经元作为先验元素.

2.3　知识插入

在获得有效知识之后, 进一步讨论如何将规则所代表的知识插入到网络当中, 以达到提升网络特征学习的目的. 知识插入网络的过程一般为利用规则对深度网络进行初始化, 这极大程度地决定着网络模型的性能[17]. 在知识插入作用下, 深度网络的初始化和训练将更加容易且有效[22]. 在网络的初始阶段就赋予一定的知识, 可以提高网络学习性能并降低对数据的依赖程度.

在特征提取部分, 置信度规则被用于初始化网络并帮助网络训练. 置信度规则的符号逻辑被用于初始化 DAE 网络结构; 置信值被用于初始化 DAE 中的权重值. 如图 3 所示, 利用一个简单的规则作为例子描述了置信度规则初始化 DAE 的过程[22].

在 DAE 被初始化之后, 对其进行自监督训练过程中, 为了保证知识能够保存在网络中而不会随着训练的进行而失效, 选择置信度较高的规则进行权值参数冻结处理. 通过这种方法既可以保证知识

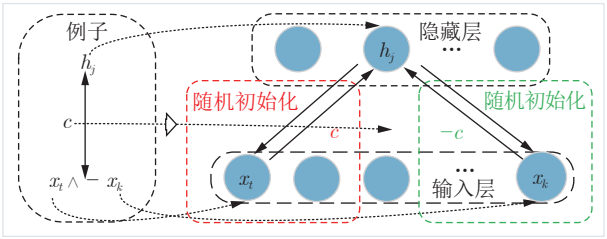


图 3　置信度规则初始化网络过程示意图

Fig. 3　The process of network initialization base on confidence rule

的有效插入, 也可以保证模型的鲁棒性. 特征提取部分具体知识插入过程如下所示:

步骤 1. 建立一个 DAE, 对每一个规则 c_j: $h_j \leftrightarrow x_1 \wedge \cdots \wedge x_n$, h_j 和 $x_1 \wedge \cdots \wedge x_n$ 分别对应目标网络 DAE 的隐藏层神经元以及输入层神经元集.

步骤 2. 确定在 h_j 与 $x_1, \cdots, x_n$ 之间的连接权重 sc_j. 如果输入神经元对应规则中的激活元素, 那么 $s = 1$, 反之则 $s = -1$. 其余的与 h_j 没有关联以及隐藏层与输出层之间的连接权重设为较小的随机值. 神经元偏差设为随机值.

步骤 3. 采用反向传播算法训练网络, 其中部分被规则初始化的连接权重不被更新. 为了保证插入的规则在训练过程中与网络较好嵌合, 利用随机数对隐藏层神经元输出进行二值化处理: 随机生成一个数值在 0 ~ 1 的随机数 R, 如果 $h_j > R$ 那么 $h_j = 1$, 反之则 $h_j = 0$.

步骤 4. 对每一个 DAE 重复步骤 1 ~ 3 进行训练, 直到所有堆叠的 DAEs 训练完成.

分类器部分仅由单层神经元构成, 所以这部分的初始化可以简化成如何将规则插入单层前向神经网络问题. 由于 MofN 规则[17, 22] 包含数和符号两部分, 故分类器的知识插入过程可以具体化为利用 MofN 规则初始化单层前向神经网络的过程.

初始化过程的主要任务是确定分类层神经元的连入权重值和偏置值. 如图 4 所示, 对一个简单的 MofN 规则:

$$\begin{aligned}&\text{IF } (w_1 \times NumTure(A_a, A_b) + \\ &\quad w_2 \times NumTure(A_c, A_d) > b_1) \text{ THEN } y = C_i\end{aligned} \tag{15}$$

首先利用其中的符号确定网络的整体结构, 其次利用 w 和 b 分别确定第 i 个分类层神经元的连入权重值和偏置, 最后添加规则中没有提到的关系并将这些权重值设为极小的随机数, 这一过程从 SDAE 的角度来看是对分类器 C 以及隐藏层 H^N 部分的初始化, 图 4 为了简洁表示省略了大部分连接线.

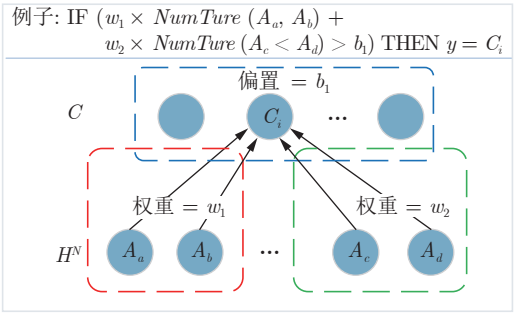


图 4 MofN 规则初始化网络过程示意图

Fig. 4 The process of network initialization based on MofN rules

随着进一步的研究发现, 将规则过多的插入分类器中反而会使网络性能降低, 这是由于网络参数被过度初始化从而使鲁棒性降低所导致的. 经过理论[32]和试验对比, 最终确定 MofN 规则的插入比率为 1/4, 其中筛选过程完全随机.

2.4 KBSDAE 训练

通过规则插入, KBSDAE 的结构参数被确定完成, 然后对网络进行进一步训练, 使其具有更好的性能. KBSDAE 的训练过程首先是进行逐个 DAE 的无监督训练, 之后进行网络微调, 但过程中的参数更新策略不同. 在自监督训练阶段, 选择将置信度关系高的参数进行冻结处理, 在训练过程中尽可能保护知识不被改变; 在微调阶段, 被 MofN 规则确定参数在更新过程中加入了抑制系数 L, 改变了这一部分参数的学习率 $\eta_r = \eta \cdot L$. 通过上述训练策略, 可以在知识插入效率和网络性能之间寻找到平衡点, 使得网络的性能被最大化提高.

在训练过程中, KBSDAE 的规则抽取和插入的乘–加操作为 11.02 KB. 这一过程消耗了一定的计算量, 但同时也加快了 KBSDAE 的收敛速度, 大幅减少了 KBSDAE 的训练耗时. 相同条件下 (训练数据 18 000 个样本), 即使加上规则抽取与插入的时间成本, KBSDAE 训练至收敛的平均训练时间仅是 SDAE 的 1.2 倍, 并且这个差距会随着数据量的增大而减小. 在预测过程中, KBSDAE 对每一例数据的乘–加操作为 4.41 KB, 内存占用为 8.33 KB. 对比深度神经网络 (如 GoogleNet[33]) 计算量更少并且内存占用量也更小, 更适合工业过程的线上识别环境.

与 SDAE 相比, KBSDAE 具有以下优点: 模型通过数据和规则两种方式进行学习, 降低了深度网络对数据的依赖性, 这在工业领域是具有重要意义的; 初始化后的网络本身具备更合理的结构参数, 使模型具备更高的识别精度和更快的收敛速度[34]. 综上所述, KBSDAE 更适合晶圆缺陷识别领域.

3 晶圆缺陷探测与识别系统

本文提出的基于 KBSDAE 晶圆缺陷识别方案如图 5 所示. 整个探测识别分为离线建模和在线探测 2 个部分. 离线建模方面, 首先对数据库中已有的晶圆图进行降噪处理突出晶圆的模式特征, 其次提取图像的几何、灰度、纹理等特征, 最后通过神经–符号系统建立缺陷探测与识别系统. 该系统第 1 步是通过正常特征数据建立基于 KBSDAE 的监控控制图, 用于晶圆缺陷探测; 第 2 步是通过缺陷特征数据构建 KBSDAE 模型, 用于晶圆缺陷识别.

3.1 图像滤噪与特征产生

晶圆图像通常参杂各种噪声, 直接使用往往不能达到预期效果, 故首先采用非线性空域滤噪技术[35]对晶圆图进行滤噪处理. 非线性空域滤噪法是直接

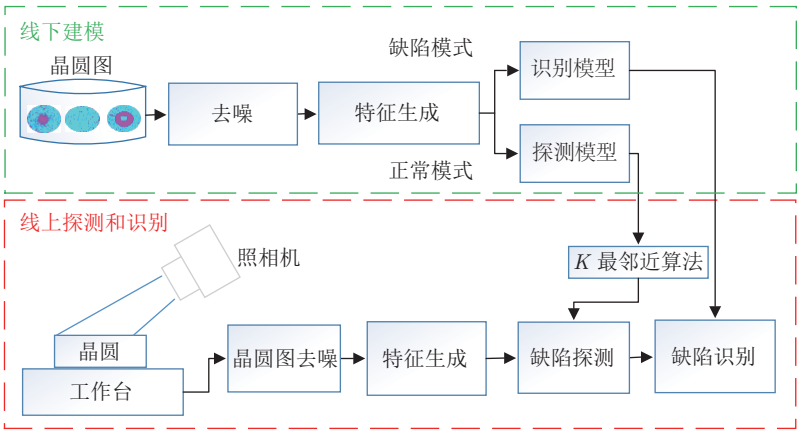


图 5 基于 KBSDAE 的晶圆表面缺陷识别系统

Fig. 5 Wafer surface defect recognition system based on KBSDAE

处理图像像素的一种滤噪方法, 本文利用像素领域内灰度值的中值代替该像素的值.

从晶圆图中直接提取有效特征可在保证模型精度的同时大大降低计算复杂度, 对本系统具有实际意义. 因此, 本文从几何、灰度、纹理、投影 4 个方面进行特征提取, 其中几何特征用于描述形状和大小, 其余特征用于描述灰度特征, 具体特征集列表如表 1 所示. 总特征维度 51 维, 其中几何特征 18 维, 投影特征 24 维, 其余特征包括重心坐标、对比度等共 9 维. 尽管提取了有效特征, 但该特征集仍具有较高维度, 并且包含很多噪音, 不适合直接输入归类器进行分类识别. 因此, 本文采用 KBSDAE 进行进一步的特征学习及分类识别.

表 1　晶圆图像特征集

Table 1　Wafer map feature set

特征类别	特征集
几何特征	区域特征、线性特征、Hu 不变矩
灰度特征	平均值、方差、歪斜度、峰值、能量、熵
纹理特征	能力、对比度、相关性、均匀度、熵
投影特征	峰值、平均幅值、均方根幅值、投影波形特性、投影峰值、投影脉冲

从晶圆中进行特征产生有以下 3 个优点: 1) 以低维的原始特征集代替高维的图像将使得深度网络模型结构更加简单有效; 2) 将图像的像素特征转换为简单的特征等可以更好地简化规则, 然后提升深度网络模型的可解释性; 3) 规则关联可理解的物理特征而不是像素特征将提高规则的可理解性与有效性.

3.2　晶圆缺陷探测与识别系统构建

整个晶圆缺陷识别过程分两步走, 首先进行缺陷探测, 其次进行缺陷识别. 缺陷探测的主要目的是区分正常和存在缺陷的晶圆. 缺陷识别的主要目的是识别晶圆缺陷的具体类别. 将缺陷探测和识别分解为 2 个问题: 1) 两分类可以有效提高故障探测性能; 2) 九分类问题转换为八分类问题, 更少的分类可有效提高深度网络模型的缺陷识别性能.

本文缺陷探测模型如图 6 上半部分所示, 主要包含基于 KBDAE 的控制图与 KBDAE 识别器两部分. 具体建模过程为: 首先利用部分数据建立并训练标准 DAEs 并利用 Confidence-RE 算法抽取置信度规则, 其次利用规则初始化基于知识的降噪自编码器 (Knowledge-based DAEs, KBDAEs) 并用另一部分数据进行训练, 最后将 KBDAEs 输出的特征数据作为控制变量建立控制图, 设定控制图信任限为 99.73% (3σ 合格率), 制造过程状态检

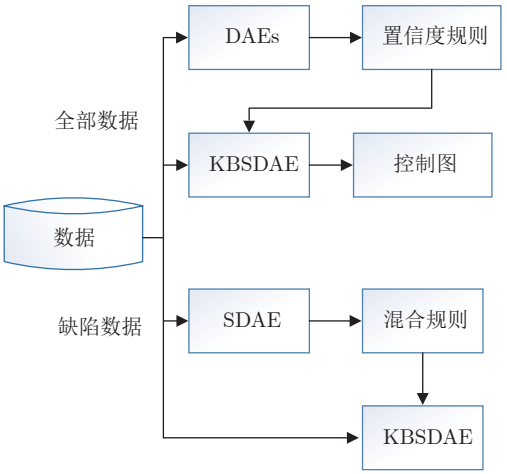


图 6　晶圆缺陷探测与识别流程

Fig. 6　The process of defect detecting and identifying on wafer

测指标为在线抽取向量特征与在控过程特征的欧氏距离 D:

$$D(x,\,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{16}$$

控制图可以在保证制造过程异常探测性能的同时, 实现制造过程状态的可视化, 是生产过程中有效的质量检测工具.

晶圆缺陷识别模型的建立过程如图 6 下半部分所示, 首先利用部分数据建立 SDAE 模型并通过规则抽取算法得到规则集 CM-R, 其次利用 CM-R 构建 KBSDAE 并用另一部分数据训练. 通过上述方法可得到一个可以被分析且具有高识别性能的 SDAE 模型.

4　晶圆缺陷探测与识别系统

WM-811K[36] 的图像数据来自实际半导体生产线. 根据晶圆图中像素位置的扫描值, 分别对正常、缺陷和空元素使用青色、品红和白色进行标注. WM-811K 数据集包含 8 个缺陷模式 (Center、Edge-ring、Edge-local、Random、Local、Scratch、Near-full、Donut) 和 None-pattern, 如图 7 所示. 数据集分为训练集和测试集, 分别用于构建模型和测试模型的性能. 用于进行故障检测和识别的晶圆片映射的详细信息如图 8 所示. 很明显, WM-811K 数据集存在类不平衡, 这将给 KBSDAE带来挑战.

4.1　晶圆表面缺陷探测

在缺陷探测系统中, 首先利用基于 KBSDAE 的监控图检测晶圆缺陷. 使用所有数据的 60% 作为

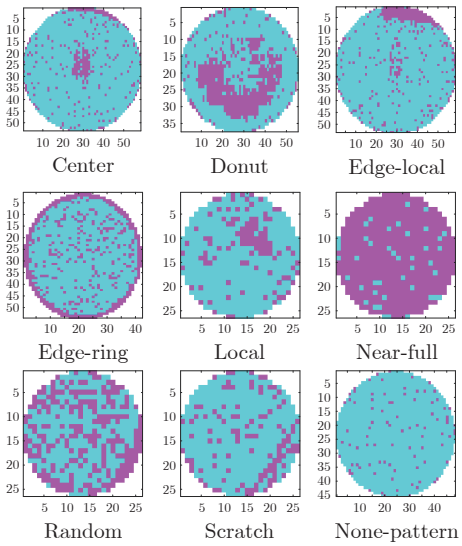


图 7 正常模式与 8 种缺陷模式的晶圆图

Fig. 7 Normal pattern and eight defect patterns of wafer

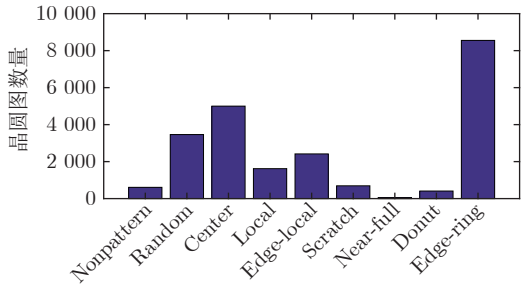


图 8 WM-811K 中晶圆图数据构成

Fig. 8 Data Structure of wafer map in WM-811K

训练集来构建 KBSDAE (其中 20% 数据用来建立标准 SDAE, 其他数据用来训练 KBSDAE), 10% 的数据作为测试集来执行缺陷检测. 为了体现 KBSDAE 的优越性, 增加了基于原始数据和 SDAE 的控制图结果进行对比. 基于原始数据、SDAE 和 KBSDAE的监控图分别如图 9 ~ 11 所示, 其中阈值设置为 99.73%, 在假报率和漏报率之间取得较好的权衡. 对比 3 个控制图可以发现 KBSDAE 控制图的表现明显优于基于原始数据和 SDAE 的控制图. 由图 11可以看出, 监控图几乎检测到了所有的缺陷, 并且不会触发太多的虚警 (虚警率为 0.05%). 结果表明, 该监测图对晶圆图缺陷的在线检测是有效的.

图 9 ~ 11 给出了基于原始数据、SDAE 和 KBSDAE 控制图的缺陷模式检出率. 表 2 给出了 3 种控制图的缺陷探测率. KBSDAE 控制图的检出率明显高于其他 2 种图, 并且不会出现对个别缺陷完全不能识别的问题. KBSDAE 控制图可以检测

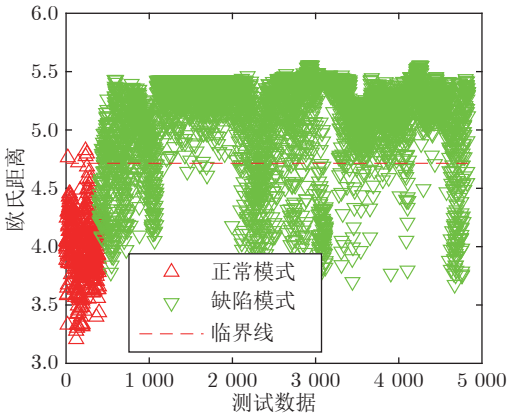


图 9 基于原始数据的控制图

Fig. 9 Control chart based on raw data

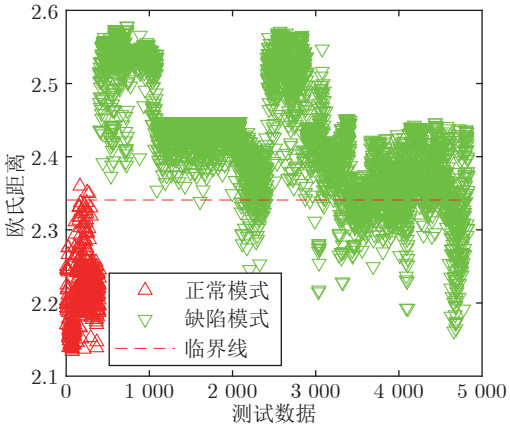


图 10 基于 SDAE 提取特征的控制图

Fig. 10 Control chart based on feature extracted by SDAE

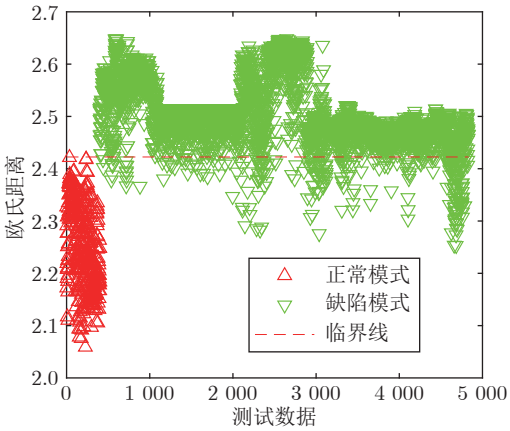


图 11 基于 KBSDAE 提取特征的控制图

Fig. 11 Control chart based on feature extracted by KBSDAE

出 93.52% 的缺陷晶圆图, 可满足工业应用的要求. 虽然 SDAE 输出特征对比原始数据更加有效, 但控

制图对个别缺陷类完全无法探测. 但是, KBSDAE 对几乎所有缺陷类可以进行有效的探测, 其缺陷探测显著优于 SDAE. KBSDAE 提取的特征可以极大地提升控制图的缺陷探测性能. 同时, KBSDAE 可以更好地处理类不平衡数据, 这是由于知识插入显著地提高了其特征学习性能.

表 2　3 种控制图的缺陷探测率 (%)
Table 2　Defect detection capabilities of three control charts (%)

模式	原始数据	SDAE	KBSDAE
Random	62.90	100	97.54
Center	99.40	99.90	97.20
Local	58.02	81.48	88.58
Edge-local	85.03	100	98.75
Scratch	99.27	98.54	86.86
Near-full	0.00	0.00	100
Donut	7.41	97.53	81.48
Edge-ring	91.10	67.19	90.86
平均值	70.89	80.58	93.52

4.2　规则有效性验证

训练 SDAE 并从容中抽取规则, 从规则的可理解性、准确度、信息保真度方面进行有效性验证. 本节从训练数据 (仅有故障数据) 中随机选取 4 000 例数据训练标准 SDAE 网络, 网络由 2 层 DAE 和全连接分类层堆叠而成, 结构为 51-60-15-8.

在验证规则可理解方面, 打印了部分 CM-R 规则并尝试通过自然语言描述其中的逻辑意义. 从 DAE 部分抽取的部分置信度规则束如表 3 所示, 规则结构与网络结构基本相似, 它们同为堆叠嵌套结构. 规则中的 x_i 代表输入层第 i 个神经元 (第 i 个维度的输入数据), h_j^l 代表第 l 个 DAE 的编码层的第 j 个神经元. 总结表 3 中的规律可知: 当输入层 x_4、x_{25}、x_{51} 尽可能大, 且 x_1、x_{21}、x_{23}、x_{50} 尽可能小时, h_{42}^1 激活的可能性较大; 当 h_{42}^1 等神经元激活, 而 h_2^1、h_{79}^1 等不激活时, h_9^2 激活的可能性较大. 这些规则很好地揭示了深度 DAE 的内部运作机理.

从全连接层抽取的 MofN 规则如表 4 所示, 描述了全连接层的推导逻辑, 并对置信度规则推导进行归纳总结. 例如: 当 $(0.68 \times x - 1.35 \times y) > 0.75$ 时, 这组数据属于第 1 类的可能性较大, 其中 x 表示 h_2^2、h_3^2、h_4^2、h_5^2、h_6^2、h_7^2、h_9^2、h_{10}^2、h_{12}^2、h_{13}^2 中被激活的神经元个数, y 表示 h_1^2、h_8^2、h_{11}^2、h_{14}^2、h_{15}^2 中被激活的神经元个数.

将表 3 和表 4 的规则结合起来, 就可以形成一套 CM-R 规则. 从表现形式和代表意义上可以得出, 这套规则有效地描述了 SDAE 网络内部结构, 达到了对深度网络进行知识抽取和网络结构解释的目的. 通过 CM-R 的表示, 神经网络中的运算逻辑可被以一种简单有效的方式进行表达. 通过对 CM-R 的推理, 规则集可以作为一个简单的分类器, 并且具备 “白盒”模型的特性. 可以通过对规则集的推导, 了解深度网络内部分类机制, 也可量化输入特征的重要程度.

可将规则集看作一种分类器, 利用 1 000 例测试数据分别对 CM-R 和 SDAE 进行准确率测试, 其中 CM-R 的准确率为 73.96%, SDAE 的准确率为 88.67%. 从测试结果可以看出规则和网络之间存在差距, 这是因为规则在提取过程中会出现信息损

表 3　部分置信度符号规则
Table 3　Part of Confidence Rule

DAE	置信度规则
DAE 1	$0.55 : h_2^1 \Leftrightarrow x_1 \wedge \neg x_2 \wedge \neg x_4 \wedge x_5 \wedge \cdots \wedge x_{21} \wedge \neg x_{22} \wedge x_{23} \wedge \neg x_{25} \wedge \cdots \wedge \neg x_{49} \wedge x_{50} \wedge \neg x_{51}$ $0.65 : h_{42}^1 \Leftrightarrow \neg x_1 \wedge \neg x_2 \wedge \neg x_3 \wedge x_4 \wedge \neg x_5 \wedge \cdots \wedge \neg x_{24} \wedge x_{25} \wedge \cdots \wedge \neg x_{49} \wedge \neg x_{50} \wedge x_{51}$ $0.56 : h_{79}^1 \Leftrightarrow x_1 \wedge x_3 \wedge \neg x_4 \wedge \cdots \wedge x_{21} \wedge x_{22} \wedge x_{23} \wedge \neg x_{24} \wedge \neg x_{25} \wedge \cdots \wedge x_{49} \wedge x_{50} \wedge x_{51}$
DAE 2	$0.72 : h_9^2 \Leftrightarrow \neg h_2^1 \wedge \neg h_5^1 \wedge h_7^1 \wedge \neg h_{10}^1 \wedge \neg h_{11}^1 \wedge \neg h_{12}^1 \wedge \cdots \wedge \neg h_{41}^1 \wedge h_{42}^1 \wedge \cdots \wedge \neg h_{77}^1 \wedge \neg h_{78}^1 \wedge \neg h_{79}^1$

表 4　部分 MofN 规则
Table 4　Part of MofN Rule

分类	MofN 规则
类别 1 ($C1$)	IF $0.68 \times NumberTure\,(h_2^2, h_3^2, h_4^2, h_5^2, h_6^2, h_7^2, h_9^2, h_{10}^2, h_{12}^2, h_{13}^2) - 1.35 \times NumberTure\,(h_1^2, h_8^2, h_{11}^2, h_{14}^2, h_{15}^2) > 0.75$ THEN $C1$
类别 4 ($C4$)	IF $3.45 \times NumberTure\,(h_5^2, h_6^2, h_7^2, h_8^2) - 0.87 \times NumberTure\,(h_1^2, h_2^2, h_3^2, h_4^2, h_9^2, h_{10}^2, h_{11}^2, h_{12}^2, h_{13}^2, h_{14}^2, h_{15}^2) > 4.73$ THEN $C4$
类别 5 ($C5$)	IF $0.85 \times NumberTure\,(h_2^2, h_4^2, h_5^2, h_6^2, h_7^2, h_8^2, h_9^2, h_{10}^2, h_{12}^2, h_{15}^2) - 1.76 \times NumberTure\,(h_1^2, h_3^2, h_{11}^2, h_{13}^2, h_{14}^2,) > 1.44$ THEN $C5$

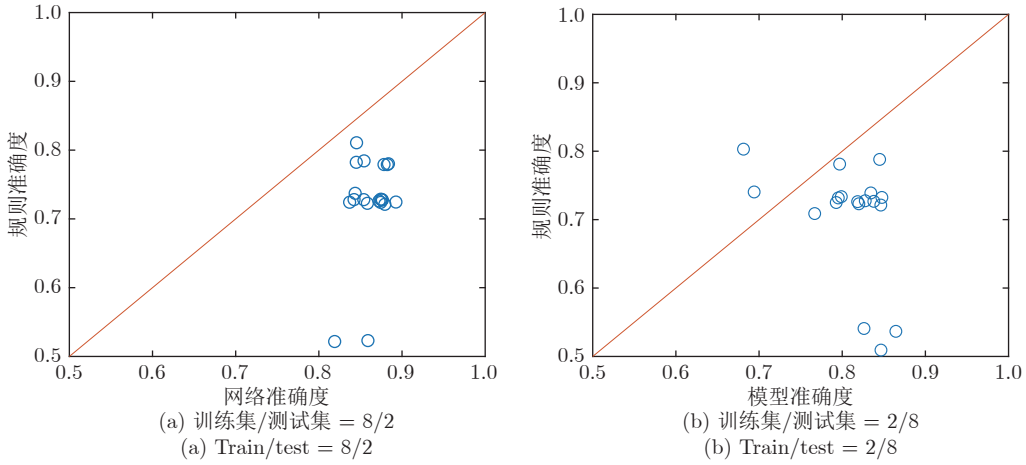
失现象. 为了验证这种信息损失对 CM-R 的影响, 对比了规则和对应标准网络在相同测试数据下的推导精度. 首先, 利用不同训练数据分别训练 20 个标准双层 DAE 网络并从中抽取规则. 其次, 对 20 个 SDAE 模型分别用 20 例不同的测试数据进行测试, 结果如图 12 所示. 图 12 横坐标表示标准网络在测试集上的预测精度, 纵坐标表示规则在测试集上的推导精度, 线代表网络和规则测试精度相同的基准线, 每个点代表一组模型 (一个标注 SDAE + 从中抽取的 CM-R) 的测试结果. 可以看出, 大部分点都在基准线附近, 证明了整套规则算法的有效性; 近乎所有点都在线下方, 证明信息损失是存在的 2 张图结果点较为密集, 证明模型具有较识别高稳定性, 即便训练数据量发生变化, 规则精度也不会发生突变. 结果表明 CM-R 规则具有较好的保真度[37]. 尽管 CM-R 规则具有一定的信息损失, 但是依然有效地提高了 KBSDAE 的特征学习性能.

4.3 KBSDAE 训练过程分析

知识插入不仅使 KBSDAE 的初始化具备了一定的模式识别能力, 而且将有效地提升 KBSDAE 的无监督训练学习和有监督的微调学习. 为了验证知识插入网络是否可以为缺陷识别带来积极影响, 首先利用规则初始化网络, 并利用余下训练数据 (仅包含缺陷数据) 训练 KBSDAE, 其次利用训练数据训练了规模相同的 SDAE. 为分析两种网络的表现, 记录了模型在无监督训练和微调阶段的均方误差变化. 由图 13 可以看出, 不管是在无监督训练还是在微调阶段, KBSDAE 的均方误差相较于 SDAE 都具有更快的收敛速度和更低的收敛区间. 这证明了利用知识初始化网络所带来的积极影响, 也进一步证明了本文提出方法的有效性.

表 5 进一步给出了 KBSDAE 在测试数据上的识别结果混淆矩阵. 这个矩阵中的对角线元素是每个缺陷模式的识别率 (总体准确率为 91.2%). 由表 5 可以看出, 大部分错误来自于对局部 (Local)、划痕缺陷 (Scratch) 和近满 (Near-full) 的错误识别, 其中 Local 和 Scratch 出现误判是由于它们本身的类别特征具有相似性导致容易混淆. Near-full 则是因为数据极少导致模型对该类的学习不足, 但在提取规则帮助下, 它被准确识别准确率达到了 84%. 图 14 是被误判的 Local 和边缘局部 (Edge-local) 的晶圆图, 它们之间存在共性, 故鉴定边界模糊容易混淆. 一般情况下, 可以接受这些错误分类的结果, 因为这些晶圆图可能同时具备一种以上模式特性. 上述结果表明, KBSDAE 在面对类不平衡数据也能对各类进行有效分类, 其主要原因是规则插入提高了 KBSDAE 的特征提取能力, 减少了数据类不平衡对网络的影响.

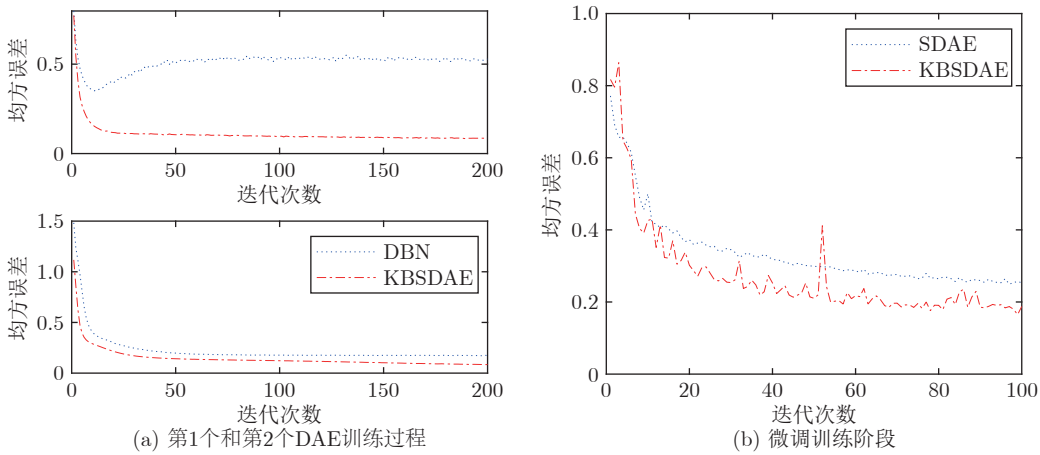
为进一步验证知识插入深度网络的优化效果, 对比了 KBSDAE 和 SDAE 在不进行微调和只进行几步微调后的测试精度. 利用相同数据分别建立了结构和训练参数相同的 SDAE 和 KBSDAE, 网络的 2 个训练阶段的学习率分别为 0.05 和 1, DAE 训练阶段噪声率为 0.05. 测试结果如图 15 所示, 可以看出, KBSDAE 在不进行微调的情况下仍具有一定的识别能力, 与 SDAE 相比提升明显. 这进一步证明了利用规则插入网络可以进一步提升 SDAE 的特征学习性能. 而经过前几步微调后的 KBSDAE 测试精度普遍高于 SDAE, 这证明了将知识代入网络可以显著提高网络的分类性能.



(a) 训练集/测试集 = 8/2
(a) Train/test = 8/2
(b) 训练集/测试集 = 2/8
(b) Train/test = 2/8

图 12 SDAE 和相应的符号规则的晶圆表面缺陷识别率对比

Fig. 12 Comparison of wafer defect recognition rates between SDAE and corresponding rules



(a) 第1个和第2个DAE训练过程
(a) The training process of the first and second DAE

(b) 微调训练阶段
(b) Fine-tuning process

图 13　KBSDAE 和 SDAE 训练过程的均方误差变化对比

Fig. 13　Comparison of mean square errors of KBSDAE and SDAE training processes

表 5　基于 KBSDAE 的晶圆缺陷识别率

Table 5　Recognition rates of defects in wafers based on KBSDAE

模式	Random	Center	Local	Edge-local	Scratch	Near-full	Donut	Edge-ring
Random	0.91	0	0.06	0	0	0	0	0.03
Center	0.01	0.99	0	0	0	0	0	0
Local	0.01	0.01	0.81	0	0.09	0	0	0.08
Edge-local	0	0.02	0	0.98	0	0	0	0
Scratch	0	0	0.03	0.02	0.83	0	0	0.12
Near-full	0	0	0.01	0	0.25	0.84	0	0
Donut	0	0	0	0.13	0	0	0.87	0
Edge-ring	0	0	0	0	0.02	0	0	0.98

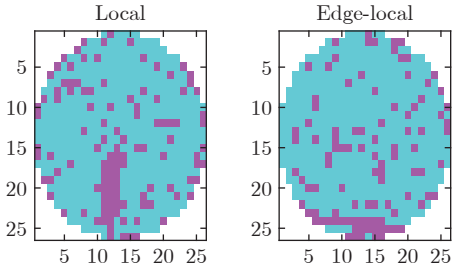


图 14　Local 和 Edge-local 模式的晶圆图

Fig. 14　Wafer maps in Local and Edge-local patterns

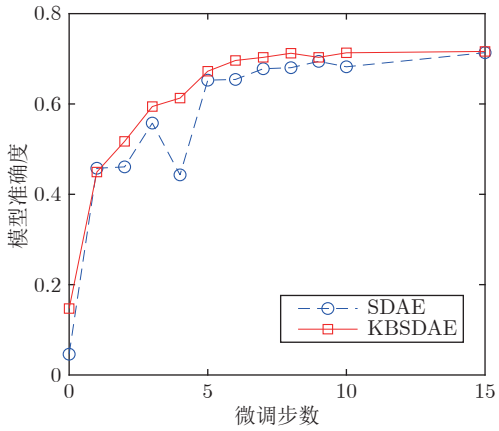


图 15　不同微调训练步数的 SDAE 与 KBSDAE 分类性能比较

Fig. 15　Comparison of classification performances between SDAE and KBSDAE with different fine-tuning steps

4.4　超参数敏感性分析

对于 KBSDAE, 网络结构、规则的插入规模等参数对其判别特征提取的有效性有显著影响. 为检验重要参数对网络识别性能的影响程度, 对网络进行参数敏感性分析. 敏感性分析是通过在一定范围内改变这些参数来实现的. 由表 6 可知, KBSDAE 的性能随着隐藏层数的增加而提高, 规则过多并不能提高 KBSDAE 的性能. 其中, 采用前 1/3 置信度规则和 1/2 分类规则构造双层 KBSDAE 时, 晶圆缺陷识别效果最好.

表 6 结构规则超参数敏感性分析

Table 6 Model hyperparameter sensitivity analysis

隐藏层数	隐节点数	置信度规则数	分类规则数	准确度 (%)
1	20 + 5	1/2	1	89.37
			1/2	88.70
			1/4	87.57
		1/3	1	89.00
			1/2	88.80
			1/4	88.57
		1/5	1	89.80
			1/2	88.97
			1/4	87.67
2	80, 15 + 5	1/2	1	86.27
			1/2	90.02
			1/4	89.00
		1/3	1	90.00
			1/2	**91.56**
			1/4	89.78
		1/5	1	90.00
			1/2	88.13
			1/4	88.90
3	80, 30, 15 + 5	1/2	1	84.23
			1/2	89.37
			1/4	89.20
		1/3	1	83.47
			1/2	87.33
			1/4	88.07
		1/5	1	84.23
			1/2	88.62
			1/4	89.05

为了检验网络模型对数据的敏感度, 对比了在不同训练数据量下 KBSDAE 和 SDAE 的识别精度. 利用相同训练数据分别训练 SDAE 和 KBSDAE, 训练数据量从 20 开始逐渐递增. 训练后的网络利用 1000 个测试数据进行识别性能测试. 结果如图 16 所示, 即使在训练数据量很小的情况下, KBSDAE 依旧具有高识别精度, 这是由于知识代入网络的结果. 并且随着训练数据量的增加, KBSDAE 识别精度也稳定高于标准 SDAE. 试验结果证明 KBSDAE 相较于 SDAE 具有更高的数据敏感度, 在缺乏训练数据的情况下依旧可以保持较高的识别精度, 这在工业应用方面是很大的提升.

4.5 结果比较

将 KBSDAE 在 WM-811K 和相关仿真数据上的分类结果与其他典型分类器进行了比较. 这些经典分类器包括 DBN、堆叠自编码器、堆叠稀疏自编码器 (Stacked sparse auto-encoder, SSAE)、SDAE、BP 神经网络 (Back propagation neural network, BPNN)、基于 KBANN 的符号神经系统 (Neuro-symbolic system for KBANN, INSS-KBANN)[38]、密集连接的卷积网络 (Densely connected convolutional network, DenseNet)[39]、残差神经网络 (Residual network, ResNet)[40]、谷歌网络 (Google inception net, GoogleNet)[33]、支持向量机–高斯核函数 (Support vector machine with Gaussian kernel, SVMG), 网络–符号的模型为符号–深度置信网络 (Symbolic-Deep belief network, SYM-DBN)[34]、局部与非局部联合线性判别分析 (Local and nonlocal preserving projection, JLNDA)[41]. 为了更加全面地测试 KBSDAE 的性能, 在本节试验中加入仿真数据[42], 这种数据被经常应用于验证模型有效性, 是根据晶圆故障的特性生成的带有噪声的数据, 同样的也具备类不平衡的缺陷. 图 17 展示了仿真数据的组成结构. DBN 和 SYM-DBN 的网络结构为 51-60-15-8, 受限玻尔兹曼机阶段的学习率和动量分别为 0.5 和 0, 微调阶段学习率为 2; SDAE 和 SSAE 的网络结构为 51-60-15-8, 学习率和动量分别为 1 和 0.5; INSS-KBANN 的网络结构为 51-60-15-8, 学习率和动量分别为 2 和 0.1; BPNN 的网络结构为 51-60-15-8, 学习率和动量分别为 2 和 0.1; DenseNet、ResNet 和 GoogleNet 都是直接识别图像的卷积神经网络模型, 所以直接利用晶圆图像数据进行训练和测试.

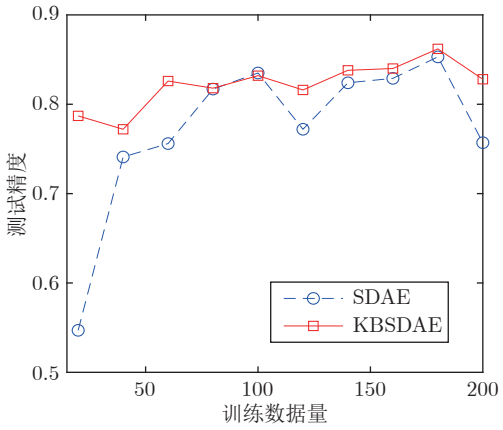


图 16 不同训练数据量下的 KBSDAE 与 SDAE 识别性能比较

Fig. 16 Comparison of classification performances between KBSDAE and SDAE with different training data volumes

对上述模型分别进行五折交叉试验, 结果如表 7 所示. 相较于传统分类器, KBSDAE 在晶圆缺陷识别上具有显著好的性能. 与直接识别图片的卷积神

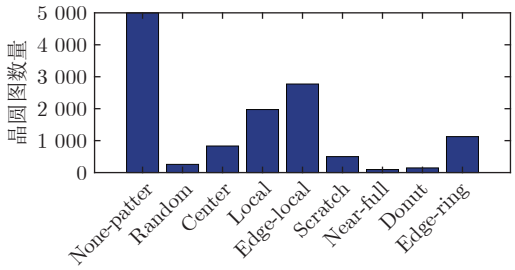


图 17　仿真数据集中晶圆图构成示意图

Fig. 17　Data structure of wafer map in simulation dataset

表 7　各种学习模型的晶圆缺陷识别率 (%)

Table 7　Wafer defect recognition rates for various learning models (%)

数据集	WM-811K	仿真
DBN	80.84	86.34
SDAE	89.87	91.28
SSAE	**86.6**	**87.96**
BPNN	80.71	89.25
DenseNet	88.6	90.69
ResNet	86.53	91.89
GoogleNet	74.32	90.63
SVMG	72.54	78.86
SYM-DBN	**85.63**	**90.58**
INSS-KBANN	81.96	92.78
JLNDA	**90.4**	**90.84**
KBSDAE	**91.14**	**95.28**

经网络模型相比, KBSDAE 的缺陷识别率更高且网络规模更小. 这是因为 KBSDAE 利用特征数据进行学习, 也说明了特征产生为 网络带来了一定的优势. 符号–神经模型 (INSS-KBANN、SYM-DBN) 相比原网络模型 (BPNN、DBN) 识别效果更好, 但需要更多时间进行知识提取与插入. 而 KBSDAE 仍然显示更好的特征学习性能. KBSDAE 在 2 种数据集上的优异表现, 也更加充分地证明了其特征学习与识别能力的优越性.

5　结束语

由于实际制造工况的复杂性, 如何解决深度神经网络在应用过程中出现的不可解释和依赖数据源的问题是晶圆缺陷识别领域迫切需要解决的问题. 本文提出了一种基于 SDAE 的神经–符号模型. 针对 SDAE 设计了适配的符号规则形式, 同时提出了适用于网络和规则的知识转化算法. 建立了一套基于 KBSDAE 的晶圆表面缺陷识别系统, 可有效地探测与识别晶圆缺陷模式. 试验结果表明, 在利用晶圆数据建模的过程中不仅规则可有效描述网络表述知识, 而且插入知识的网络同时具备高识别性能. 在未来研究中, 将继续探索神经–符号系统, 尝试更复杂深度网络模型 (比如卷积神经网络), 提高模型性能和可解释性.

References

1　Hansen M H, Nair V N, Friedman D J. Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects. *Technometrics*, 1997, **39**(3): 241−253

2　Hsieh K L, Tong L I, Wang M C. The application of control chart for defects and defect clustering in IC manufacturing based on fuzzy theory. *Expert Systems With Applications*, 2007, **32**(3): 765−776

3　Hess C, Weiland L H. Extraction of wafer-level defect density distributions to improve yield prediction. *IEEE Transactions on Semiconductor Manufacturing*, 1999, **2**(2): 175−183

4　Friedman D J, Hansen M H, Nair V N, James D A. Model-free estimation of defect clustering in integrated circuit fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 1997, **10**(3): 344−359

5　Yuan T, Kuo W. Spatial defect pattern recognition on semiconductor wafers using model-based clustering and Bayesian inference. *European Journal of Operational Research*, 2008, **190**(1): 228−240

6　Yu Jian-Bo, Lu Xiao-Lei, Zong We-Zhou. Wafer defect detection and recognition based on local and nonlocal linear discriminant analysis and dynamic ensemble of Gaussian mixture models. *Acta Automatic sinica*, 2016, **42**(1): 47−59
(余建波, 卢笑蕾, 宗卫周. 基于局部与非局部线性判别分析和高斯混合模型动态集成的晶圆表面缺陷探测与识别. 自动化学报, 2016, **42**(1): 47−59)

7　Huang C J. Clustered defect detection of high quality chips using self-supervised multi-layer perceptron. *Expert Systems With Applications*, 2007, **33**(4): 996−1003

8　Baly R, Hajj H. Wafer classification using support vector machines, *IEEE Transactions on Semiconductor Manufacturing*, 2012, **25**(3): 373−383

9　Xie L, Huang R, Gu N, Zhi C. A novel defect detection and identification method in optical inspection. *Neural Computing and Applications*, 2014, **24**(7−8): 1953−1962

10　Chao L C, Tong L I. Wafer defect pattern recognition by multi-class support vector machines by using a novel defect cluster index. *Expert Systems With Applications*, 2009, **36**(6): 10158−10167

11　Nakazawa T, Kulkarni D V. Wafer map defect pattern classification and image retrieval using convolutional neural network. *IEEE Transactions on Semiconductor Manufacturing*, 2018, **31**(2): 309−314

12　Fang Xin, Shi Zheng. Wafer defect detection and classification algorithms based on convolutional neural network. *Computer Engineering*, 2018, **44**(8): 218−223
(邡鑫, 史峥. 基于卷积神经网络的晶圆缺陷检测与分类算法. 计算机工程, 2018, **44**(8): 218−223)

13　Yu J B. Enhanced stacked denoising utoencoder-based feature learning for recognition of wafer map defects. *IEEE Transactions on Semiconductor Manufacturing*, 2019, **32**(4): 613−624

14　Lee H, Kim Y, Kim C O. A deep learning model for robust wafer fault monitoring with sensor measurement noise. *in IEEE Transactions on Semiconductor Manufacturing*, 2017, **30**(1): 23−31

15　Sun Chen, Zhou Zhi-Hua, Chen Zhao-Qian. Study on rule extraction of neural network. *Application Research of Computers*, 2000, **17**(2): 34−37
(孙晨, 周志华, 陈兆乾. 神经网络规则抽取研究. 计算机应用研究,

2000, **17**(2): 34−37)

16 Gallant S I. Connectionist expert systems. *Communications of the ACM*, 1988, **31**(2): 152−169

17 Towell G G, Shavlik J W. Knowledge-based artificial neural networks. *Artificial Intelligence*, 1994, **70**(1−2): 119−165

18 Garcez A S A, Zaverucha G. The connectionist inductive learning and logic programming system. *Applied Intelligence*, 1999, **11**(1): 59−77

19 Garcez A A, Gori M, Lamb L C, Serafini L. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. arXiv preprint, 2019, arXiv: 1905.06088

20 Odense S, Garcez A A. Extracting m of n rules from restricted boltzmann machines. In: Proceedings of the 2017 International Conference on Artificial Neural Networks. Alghero, Italy: Springer, 2017. 120−127

21 Tran S N, Garcez A S D. Deep logic networks: Inserting and extracting knowledge from deep belief networks. *IEEE Transactions on Neural Networks & Learning Systems*, 2018, **29**(2): 246−258

22 Liu Guo-Liang, Yu Jian-Bo. Knowledge based stacked denoising autoencoder. *Acta automatic sinica*, 2022, **48**(3): 774−786
(刘国梁, 余建波. 知识堆叠降噪自编码器. 自动化学报, 2022, **48**(3): 774−786)

23 Hitzler P, Bianchi F, Ebrahimi M, Sarker M K. Neural-symbolic integration and the semantic web. *Sprachwissenschaft*, 2020, **11**(1): 3−11

24 Bennetot A, Laurent J L, Chatila R, Díaz-Rodríguez N. Towards explainable neural-symbolic visual reasoning. ArXiv Preprint, 2019, ArXiv: 1909.09065

25 Li S, Xu H, Lu Z. Generalize symbolic knowledge with neural rule engine. arXiv preprint, 2018, arXiv: 1808.10326v1

26 Sukhbaatar S, Szlam A, Weston J, & Fergus R. End-To-End memory networks. ArXiv Preprint, 2015, ArXiv: 1503.08895

27 Sawant U, Garg S, Chakrabarti S, Ramakrishnan G. Neural architecture for question answering using a knowledge graph and web corpus. *Information Retrieval Journal*, 2019, **22**(3−4): 324−349

28 Liang C, Berant J, Le Q, Forbus K D, Ni L. Neural symbolic machines: Learning semantic parsers on freebase with weak supervision. arXiv preprint, 2016, arXiv: 1611.00020

29 Salha G, Hennequin R, & Vazirgiannis M. Keep it simple: Graph autoencoders without graph convolutional networks. arXiv preprint, 2019, arXiv: 1910.00942

30 Salha G, Hennequin R, Tran V A, & Vazirgiannis M. A degeneracy framework for scalable graph autoencoders. arXiv preprint, 2019, arXiv: 1902.08813

31 Vincent P, Larochelle H, Bengio Y, Manzagol P A. Extracting and composing robust features with denoising autoen-coders. In: Proceedings of the 25th International Conferenceon Machine Learning. Helsinki, Finland: 2008. 1096−1103

32 Towell G G, Shavlik J W. Extracting refined rules from knowledge-based neural networks. *Machine learning*, 1993, **13**(1): 71−101

33 Szegedy C, Liu W, Jia Y, Sermanet P, Rabinovich A. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington DC, USA: 2015. 1–9

34 Tran S N, Garcez A S D. Knowledge extraction from deep Belief networks for images. In: Proceedings of the 2013 IJCAI-Workshop Neural-Symbolic Learn. Washington DC, USA: 2013. 1–6

35 Wang C H, Kuo W, Bensmail H. Detection and classification of defect patterns on semiconductor wafers. *IIE Trans-Actions*, 2006, **38**(12): 1059−1068

36 Wu M J, Jang J S R, Chen J L. Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *IEEE Transactions on Semiconductor Manufacturing*, 2015, **28**(1): 1−12

37 Valiant L G. Three problems in computer science. *Journal of the ACM*, 2003, **50**(1): 96−99

38 Fernando S O, Amy B. INSS: A hybrid system for constructive machine learning. *Neural Computing*, 1999, **28**(1−3): 191−205

39 Huang G, Liu Z, Maaten L V D, Weinberger K Q. Densely connected convolutional networks. In: Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, Hawaii, USA: 2017. 2261–2269

40 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: 2016. 770−778

41 Yu J, Lu X. Wafer map defect detection and recognition using joint local and nonlocal linear discriminant analysis. *IEEE Transactions on Semiconductor Manufacturing*, 2016, **29**(1): 33−43

42 Di Palma F, De Nicolao G, Miraglia G, Pasquinetti E, Piccinini F. Unsupervised spatial pattern classification of electrical-wafer-sorting maps in semiconductor manufacturing. *Pattern Recognition Letters*, 2005, **26**(12): 1857−1865

刘国梁 同济大学机械与能源工程学院硕士研究生. 2018 年获上海大学学士学位. 主要研究方向为机器学习, 深度学习和智能质量管控.
E-mail: guoliangliutt@163.com
(**LIU Guo-Liang** Master student at the School of Mechanical and Energy Engineering, Tongji University. He received his bachelor degree from Shanghai University in 2018. His research interest covers machine learning, deep learning and intelligent quality control.)

余建波 同济大学机械与能源工程学院教授. 2009 年获上海交通大学博士学位. 主要研究方向为机器学习, 深度学习, 智能质量管控, 过程控制, 视觉检测与识别. 本文通信作者.
E-mail: jbyu@tongji.edu.cn
(**YU Jian-Bo** Professor at the School of Mechanical and Energy Engineering, Tongji University. He received his Ph.D. degree from Shanghai Jiaotong University in 2009. His research interest covers machine learning, deep learning, intelligent quality control, process control, visual inspection and identification. Corresponding author of this paper.)