

特集号招待論文

# ビッグデータを活用した歩留解析支援システム“歩留新聞”による解析作業時間の短縮

中田 康太<sup>1</sup> 高木 健太郎<sup>1</sup> 陶 亜玲<sup>1</sup>

<sup>1</sup> (株) 東芝

半導体製造における歩留解析では、製品の品質検査結果と各工程の処理履歴から不良の原因を特定し、対策に繋げることで生産性を向上している。半導体の製造プロセスから得られるデータは大量かつ複雑であるため、人手による歩留解析では作業に時間がかかることが問題となっている。本稿では、機械学習・データマイニングの技術を用いて不良の発生状況の可視化と不良原因装置の推定を網羅的に行う歩留解析支援システム「歩留新聞」について紹介する。歩留新聞はウェハ上の特徴的な不良の出現パターン（不良マップ）を自動で分類し、それぞれの原因装置候補を抽出する。不良マップの分類結果と原因装置候補を技術者に提示することで、不良1件あたりの解析時間を平均6時間から2時間に短縮した。ここでは、歩留新聞の概要とともに、コア技術となる深層学習技術について紹介し、製造現場における機械学習技術適用の課題とその解決方法について議論する。

## 1. はじめに

最新の半導体工場では大量の製造データが収集され、生産性の向上に利用されている。生産性向上の取り組みの1つである歩留解析では、技術者が製品の品質検査結果と各工程の処理履歴から不良の原因を特定し、対策を行うことで製品の歩留を向上している。大量のデータを用いることで、過去には不可能だった詳細な歩留解析が可能になり、大幅な生産性向上が期待されている。一方で、データは膨大かつ複雑になっており、技術者が生産性向上に繋がる有用な情報を読み取ることはますます困難になっている。

図1は、歩留解析の課題の概要を表している。大規模な半導体工場では数十から数百の種類の製品が生産される。それぞれの製品で品質検査が行われ、各製品チップに対して数十から数百の検査項目で不良の有無が判定される。製品チップが不良と判定された場合には、その検査項目が不良カテゴリとして付与される。半導体の製品はウェハと呼ばれる円盤上に製造され、製造プロセスで問題が起こった場合にはウェハ上に特徴的な不良パターンが出現することが知られている。本稿では、ウェハ上の不良パターンを不良マップと呼ぶ。

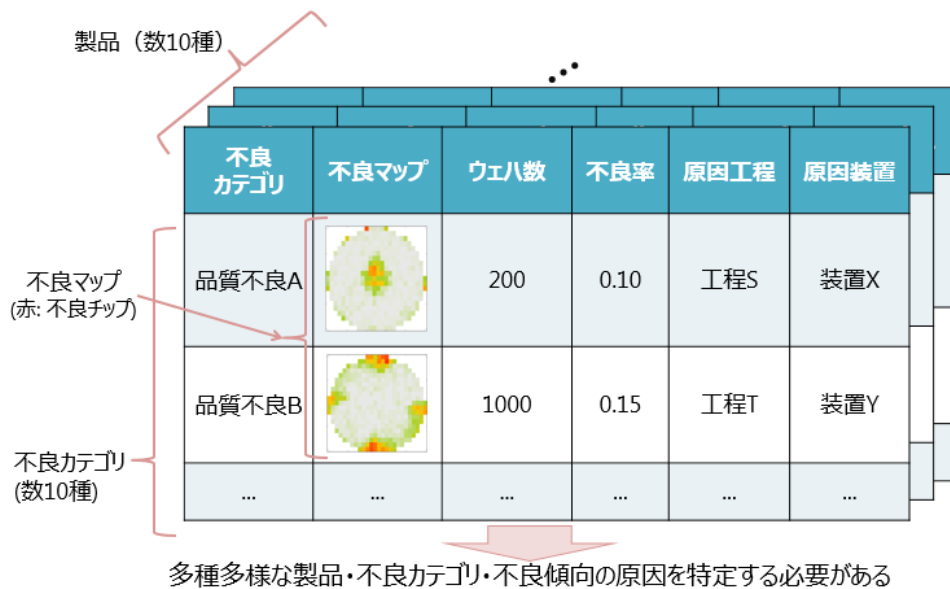


図1 半導体工場における歩留解析の課題の概要

類似した不良マップを持つウェハが複数発生した場合には、製造プロセスにその不良マップを引き起こした共通の原因があると仮定して、原因を調査する。不良マップは製品によって多様であり、過去にはない新たな不良マップが発生することもある。製品、不良カテゴリ、不良マップの組み合わせは膨大になるため、技術者がその中から早期に重要な不良を発見して原因の特定を行うことが、生産性向上への大きな課題の1つとなる。

この課題に対し、我々は機械学習・データマイニングの技術を用いて、不良マップの発生状況の可視化と不良原因装置の推定を網羅的に行う歩留解析支援システム「歩留新聞」を開発した。歩留新聞は、ウェハ上の特徴的な不良マップを自動で分類し、それぞれの原因装置候補を抽出する。また、解析の完了した不良マップを既知の不良としてシステムに学習させることで、不良の再発を自動で監視する。本稿では、歩留新聞の概要と、歩留新聞を実現する機械学習技術を紹介する。

## 2. 歩留解析支援システム「歩留新聞」

我々は、機械学習・データマイニングの技術を組み込んだ歩留解析支援システム「歩留新聞」を開発した ([1], [10]) .

図2は、不良マップ分類と原因装置推定を用いた歩留新聞のインターフェースの例を表している。図2では、技術者は不良マップ、発生枚数、発生トレンド、原因装置候補のサマ리를人手の作業なしに確認することができる。技術者が新たな不良を発見した場合には、原因装置の候補から詳細な調査を行うことが可能であり、歩留解析作業にかかる時間を大幅に短縮することができる。教師なし学習のクラスタリング手法とパターンマイニング手法を活用することで、全製品・全不良に対してサマ리의出力が可能になっている。

製品	不良カテゴリ	不良パターン	発生トレンド	ウェハ数	原因候補
製品P	カテゴリ1		 時間	300	1 工程 X: 装置 X1 2 工程 Y: 装置 Y4 3 工程 Z: 装置 Z9
製品Q	カテゴリ2		 時間	200	1 工程 A: 装置 A5 2 工程 B: 装置 B2 3 工程 C: 装置 C4
製品R	カテゴリ3		 時間	100	1 工程 D: 装置 D4 2 工程 E: 装置 E2 3 工程 F: 装置 F7

図2 歩留新聞のインターフェースの例

図3は、深層学習手法を用いた既知不良マップの再発監視のインターフェースの例を表している。図3では、技術者は注目する不良マップについて、点線内のウェハのデータを用いて学習を行っている。縦軸はウェハの枚数、横軸は時間を表しており、赤点は注目する不良が起こっていると判定されたウェハの1日あたりの枚数、緑線は全体のウェハ数を表している。図3のインターフェースにより、技術者は長期の発生トレンドを容易に監視し、不良の再発に迅速に対応することができる。

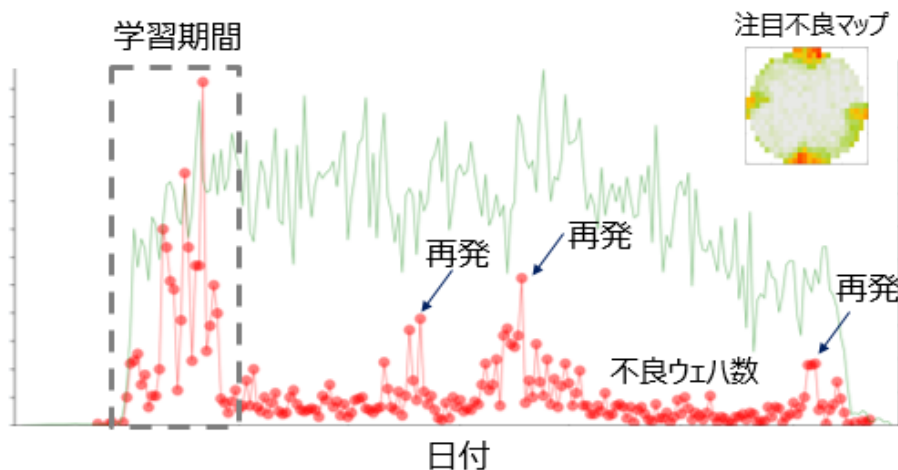


図3 歩留新聞による既知不良マップの再発監視

### 3. 歩留新聞を実現する機械学習・データマイニング技術の概要

#### 3.1 不良マップの自動分類

歩留新聞では、不良マップを自動分類するためにクラスタリング手法を導入している。クラスタリングはデータ間の距離や類似度に応じてデータをクラスタと呼ばれるグループに分割する手法で、人手による作業が必要ないメリットもあり、さまざまな分野で利用されている。

クラスタリング手法を用いて不良マップを自動分類するアプローチが過去に提案されている。たとえば、文献[2], [3]では、K-Means法 ([4]) といった代表的なクラスタリング手法を用いて不良マップを分類している。ここでは、各クラスタに属する不良マップの数が、その不良マップ

の発生枚数を表すことになる。

クラスタリング手法は、人手による教師ラベルが必要ない点が大きなメリットになる。図4は、クラスタリング手法による不良マップの自動分類の概要を表している。ここでは、類似した不良マップを分類して、それぞれのクラスタの平均的な不良マップを可視化することで、1枚1枚のウェハを技術者が目視で確認することなく、不良マップのパターンと発生枚数を把握することができる。

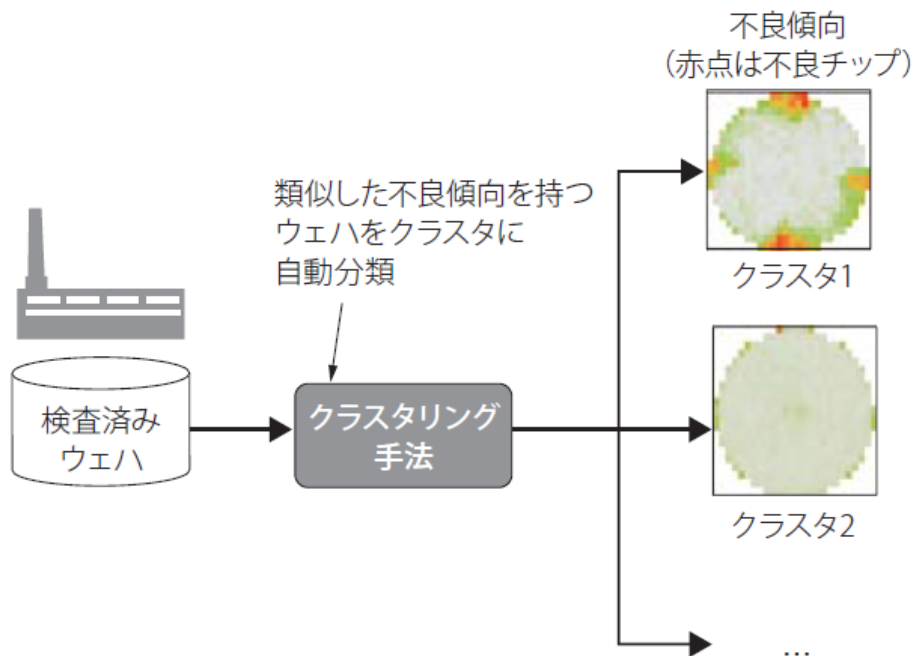


図4 クラスタリングによるウェハの不良マップの自動分類

大規模な半導体工場では日々大量の製品が生産されているため、クラスタリングの処理速度も重要な要素になる。歩留新聞では、全製品・全不良カテゴリのクラスタリングを完了するために、Sparkによる並列分散処理基盤を導入した。32並列の計算ユニット上でクラスタリングを行うことで、従来のクラスタリングの処理速度を最大で72.5倍高速化した。この高速化により、全製品・全不良カテゴリに対して数時間で処理を完了することができ、前日までの不良の発生状況を毎朝確認することが可能になる。

クラスタリングによる不良マップ分類では、データのインバランス性による精度低下の解決が課題となる。データのインバランス性とは、一部のクラスのデータがほかのクラスに比べて少量なデータを表す。インバランス性が高いデータでは、機械学習の精度が著しく低下することが知られている。

一般的に、不良の発生頻度は良品と比較して小さいため、自動分類の精度低下が懸念される。この課題に対処するため、我々はインバランスデータに対しても頑健な深層クラスタリング手法を開発した。このクラスタリング手法の詳細について第4章で紹介する。

### 3.2 不良原因装置推定

各不良マップに対する原因装置を自動推定するためにパターンマイニング手法を導入した。図5は、原因装置の自動推定に用いるデータとパターンマイニング手法の概要を表している。パターンマイニング手法の代表的な応用として、顧客の買い物履歴（トランザクション）から、同時に購入される商品のパターンを抽出する例が知られている。ここでは、ウェハが通ってきた各工程の処理装置履歴をトランザクションと見なして、不良マップに固有の装置の頻出パターンを抽出することで、原因装置の推定を行う。

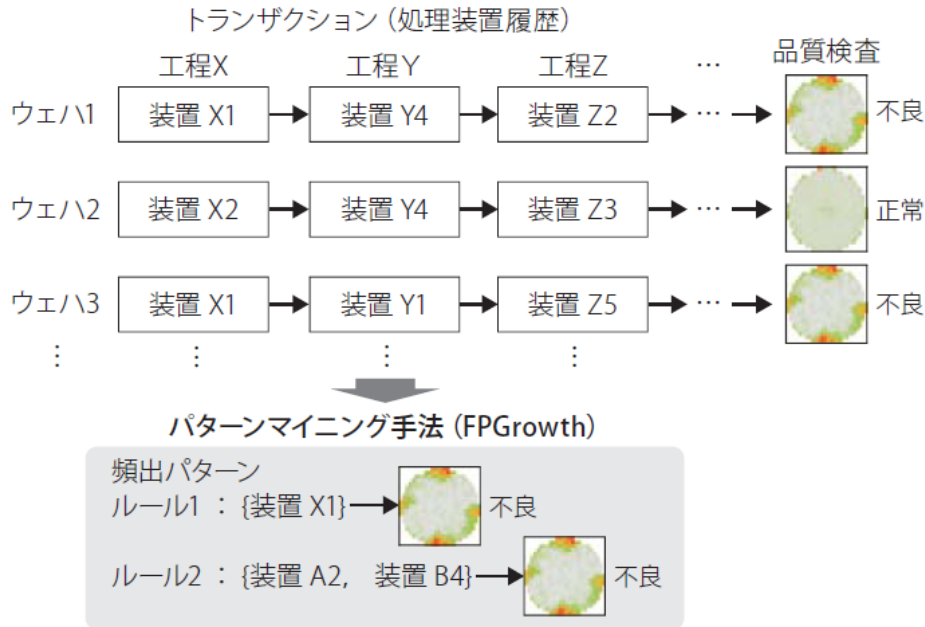


図5 パターンマイニング手法による装置抽出

図6は、歩留新聞の原因装置推定方法の概要を表している。図6では、全ウェハの処理装置履歴と、不良マップが発生しているウェハの処理装置履歴からそれぞれ装置の頻出パターンを抽出する。さらにそれぞれのパターン間でカイ2乗検定を行うことで、不良傾向が発生しているウェハに特有の装置パターンをランキング形式で表示する。

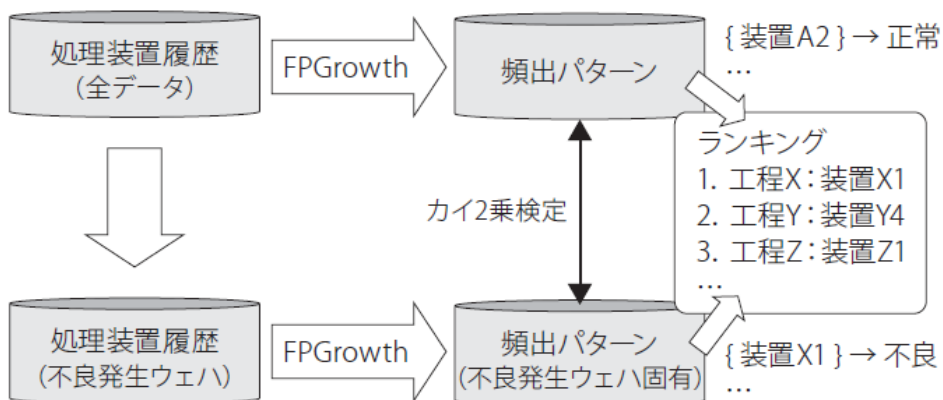


図6 原因装置の推定方法

ここではパターンマイニング手法としてFPGrowthを利用する ([5], [6]) . FPGrowthは大規模なデータから効率的に頻出パターンを抽出することができ、高速で分散処理が可能なことから、全製品・全不良カテゴリ・全不良傾向に対して原因装置推定を行うことができる。

本手法により、実際に発生した4つの不良事例について真の原因の装置を3位以内に提示できることを確認した。現在は、技術者からのフィードバックをもとに、継続的に原因推定の評価と精度向上に取り組んでいる。

### 3.3 既知不良マップの監視

図7は、歩留新聞における既知不良マップの監視の概要を表している。ここでは、標準的な教師あり学習のアプローチにより、既知の不良マップの再発監視を行う ([7], [8], [9]) . 教師あり学習では、技術者がウェハに対して注目する不良の発生有無を教師ラベルとして付与し、教師ラベルから学習した分類モデルを用いて新たなウェハを自動分類し、再発を検知する。

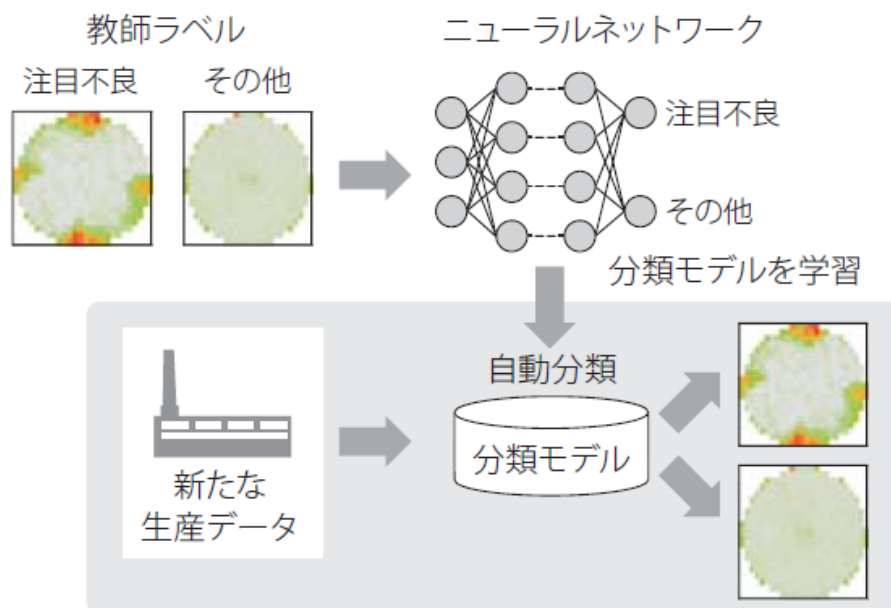


図7 既知の不良マップの再発監視

本稿では、既知不良マップの学習に深層学習手法を用いる。深層学習手法は画像認識を始めとして多くの応用分野で成功を収めている。ウェハの不良マップはシンプルな画像と見なすことができるため、深層学習手法により高性能の自動分類が期待できる。

深層学習手法により高い性能を得るためには、ネットワーク構造を含む多くのパラメータを適切に設定する必要があることが知られている。本稿では、第5章において、実データを用いた実験を行い、特に性能に対して影響の大きかったネットワーク構造とパラメータに関して考察を行う。

---

## 4. 教師なしで少量の不良を高精度に分類する深層クラスタリング技術

---

### 4.1 深層クラスタリングRDEC

歩留新聞では、クラスタリング手法により類似した不良マップを自動で分類して不良の発生状況を可視化する。第3章で述べたように、クラスタリングによる不良マップ分類では、データのインバランス性による精度低下の解決が課題となる。本稿では、少量の不良マップを正確に分類する性質を持つ独自のクラスタリング手法Regularized Deep Embedded Clustering (RDEC) について紹介する[11].

近年では、深層学習をベースにしたクラスタリング手法が提案されている。従来では、深層学習をクラスタリングに用いる場合には、たとえば自己符号化器により特徴量抽出を行い、その特徴量空間でクラスタリングを行う2段階の手法が提案されていた[12]. 2016年以降は、深層学習によりクラスタリングに好適な潜在空間を学習する手法が複数提案されており、手書き数字認識などのタスクで過去の分類精度を凌駕する性能が報告されている (e.g.[13], [14], [15], [16]) .

RDECは、深層クラスタリング手法Deep Embedded Clustering (DEC, [13]) のロス関数に、Virtual Adversarial Training (VAT, [17]) による正則化項を導入することで、分類性能を向上している。

図8を用いてRDECの概要を説明する。深層クラスタリング手法DECは初期の潜在空間におけるクラスタ中心の初期値を決定し、そのクラスタ中心にデータが凝縮されるようにネットワークの重みを更新する。クラスタリングではクラスタ中心にデータが集まり、クラスタ間の距離が大きいほど分類が容易になることから、DECはクラスタリングに好適な潜在空間を学習していると考えることができる。

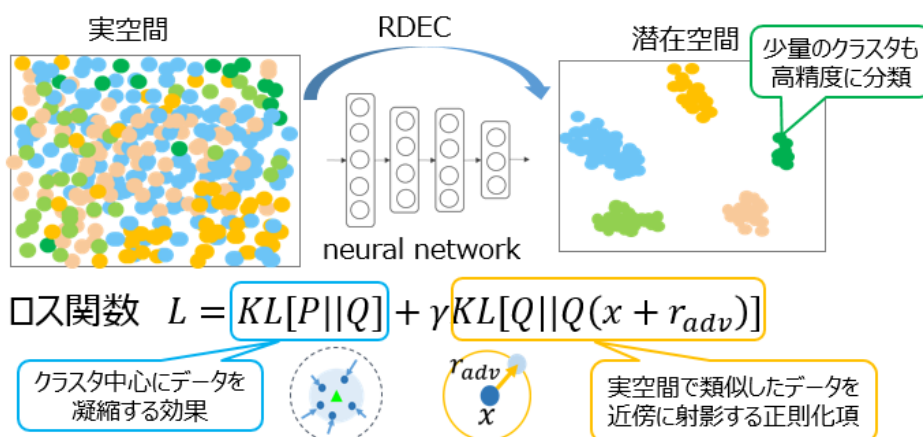


図8 深層クラスタリング手法RDECの概要

図8のロス関数の第1項は、DECの損失関数と同一で、確率分布QとPのKullback-Leibler距離を表している。確率分布Qはt分布を用いて以下の式で与えられる。

$$q_{ij} = \frac{(1 + \|z_i - u_j\|^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{j'} (1 + \|z_i - u_{j'}\|^2/\alpha)^{-\frac{\alpha+1}{2}}}$$

ここで $q_{ij}$ はデータ $i$ がクラスタ $j$ に属する確率を表しており、 $z_i$ は潜在空間におけるデータ $i$ のベクトル、 $u_j$ は潜在空間におけるクラスタ $j$ のクラスタ中心のベクトル、 $\alpha$ は定数を表している。この式では、潜在空間においてデータ $z_i$ がクラスタ中心 $u_j$ に近いほど確率 $q$ の値が高くなる。確率分布 $P$ は、 $q_{ij}$ を用いて以下の式で表される。

$$p_{ij} = \frac{q_{ij}^s / f_j}{\sum_{j'} q_{ij'}^s / f_{j'}}$$

$$f_j = \sum_i q_{ij}$$

ここで、 $s$ は定数であり、1より大きな値が用いられる。確率 $q$ は潜在空間でのクラスタ中心への距離に応じてそのデータがクラスタに属する確率が算出されるが、分布 $P$ では、その $q$ の値を $s$ 乗することで、クラスタ中心に近いデータの確率値がより高くなり、直感的には元の分布が凝縮された分布になる。

DECのロス関数ではこの分布 $P$ と $Q$ のKullback-Leibler距離が小さくなるように学習が進むため、データが最も近いクラスタ中心に凝縮される潜在空間が学習されると解釈することができる。また、どのクラスタ中心からも距離が遠く、 $q_{ij}$ の値が小さいデータについては、 $f_j$ が小さいクラスタ $j$ に属するように学習が進む。 $f_j$ が小さいクラスタは属するデータ数が少ないことを意味しているため、第1項では、どのクラスタ中心からも遠いデータはデータ数が少ないクラスタに属するように潜在空間が学習される。

図8のロス関数の第2項は、RDECで導入した正則化項を表している。第2項はデータ $x$ が潜在空間で最も急激に変化する方向に微小な摂動 $r_{adv}$ を加え、元のデータと摂動を加えたデータの潜在空間でのKullback-Leibler距離が小さくなるようにニューラルネットワークの重みを学習する。以下の式は、摂動 $r_{adv}$ の決定方法を表している。

$$r_{adv} = \arg \max_{r; \|r\| \leq \epsilon} KL[Q, Q(x+r)]$$

第2項は、VATと呼ばれる手法に基づいている[17]。VATはデータの実空間での近傍が潜在空間でも近傍になるように学習を行う手法である。ここでは、VATは実空間でデータ群が近傍に密集している場合に、それらのデータ群が潜在空間でも近傍に写像されるようネットワークの重みを学習する効果が期待される。

図9は、RDECのロス関数の効果の概要を表している。図9では、手書き数字の0と6に対して、0が少量で6が多量のインバランスデータを想定し、データを2つのクラスタに分類している。DECでは、データとクラスタ中心の距離を用いてクラスタが凝縮するように潜在空間を学習するため、少量のクラスタ境界にあるデータは小さいクラスタに凝縮され、少量の0クラスタに6が混入してしまう。



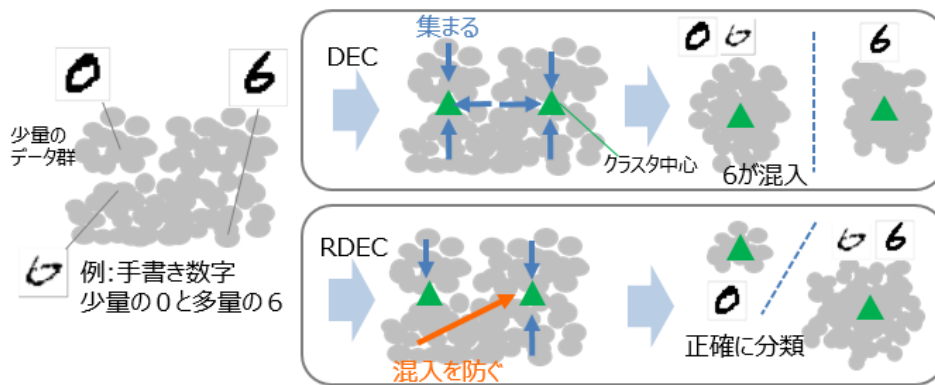


図9 RDECのロス関数の効果の概要

RDECでは、第2項を導入することで実空間の近傍のデータは潜在空間でも近傍に写像され固まって動くため、データの混入が抑制され、クラスタの大きさが異なる場合でも正確に分類を行うことができる。

#### 4.2 深層クラスタリングに関する性能実験

本節では、公開データと実データを用いた精度実験における、RDECについて分類性能を示す。

本節では、公開データセットとして手書き数字（MNIST）、ニュース記事（Reuters）、一般画像（STL）を用いる。これらの3つのデータは多くの論文でベンチマークとして用いられている。

また、3つのデータセットについて人為的に特定のクラスの割合を0.1に減少させ、インバランス性を持たせたデータセットMNIST\_Imb, Reuter\_Imb, STL\_Imbを作成した。

実データセットとして、半導体工場で収集した不良マップデータセットを用いる。不良マップは、最も少ないクラスの割合はすべてタの0.6%となっており、非常に極端なインバランス性を持っている。実験に用いるデータは、文献[11]に詳しい。

表1は、クラスタリング手法の各データセットに対する分類精度を表している。ここでは、ベースラインとしてKmeans法、自己符号化器（AE）による特徴量抽出とKmeansを組み合わせたAE+Kmeansの性能も併せて示している。ネットワーク構造はすべて共通とし、全結合層3層で潜在空間の次元数（ノード数）を10としている。

表1 クラスタリング手法と分類精度

	クラス	Kmeans	AE +Kmeans	DEC	RDEC
MNIST	10	53.38	89.93	93.98	<b>98.41</b>
MNIST_Imb	10	50.44	86.52	88.11	<b>94.82</b>
Reuter	4	53.33	74.17	73.88	<b>75.45</b>
Reuter_Imb	4	55.58	72.89	71.85	<b>73.87</b>
STL	10	72.66	77.93	84.37	<b>86.59</b>
STL_Imb	10	76.70	82.08	87.21	<b>88.63</b>
不良マップ	11	38.00	41.95	42.10	<b>79.86</b>

表1では、RDECにより公開データセットに対して精度の向上が確認できる。MNISTでは、DECの導入により分類精度が53.38%から93.98%に向上しているが、RDECではDECを上回る98.41%の精度が得られており、世界最高レベルの性能が得られている。

インバランス性の高いデータセットに対しては、RDECとそのほかの手法で性能の差が大きくなっている。不良マップデータセットは、現実の半導体工場で発生する非常にインバランスな不良を分類するタスクであるが、RDECはこのデータセットに対しても79.86%の高い分類性能を実現している。

RDECは、ロス関数の2つの項を効果的に組み合わせることで高い精度を達成している。図10は、RDECのロス関数のパラメータ $\gamma$ の依存性を示している。パラメータ $\gamma$ は、第1項と第2項のバランスを表しており、 $\gamma$ が0のときにはDECのロス関数と一致する。横軸は $\gamma$ の値、縦軸は青線が通常のMNIST、緑線がMNIST\_Imbの分類精度をエラーバー付きで表している。エラーバーは5回の実験の標準偏差を表している。

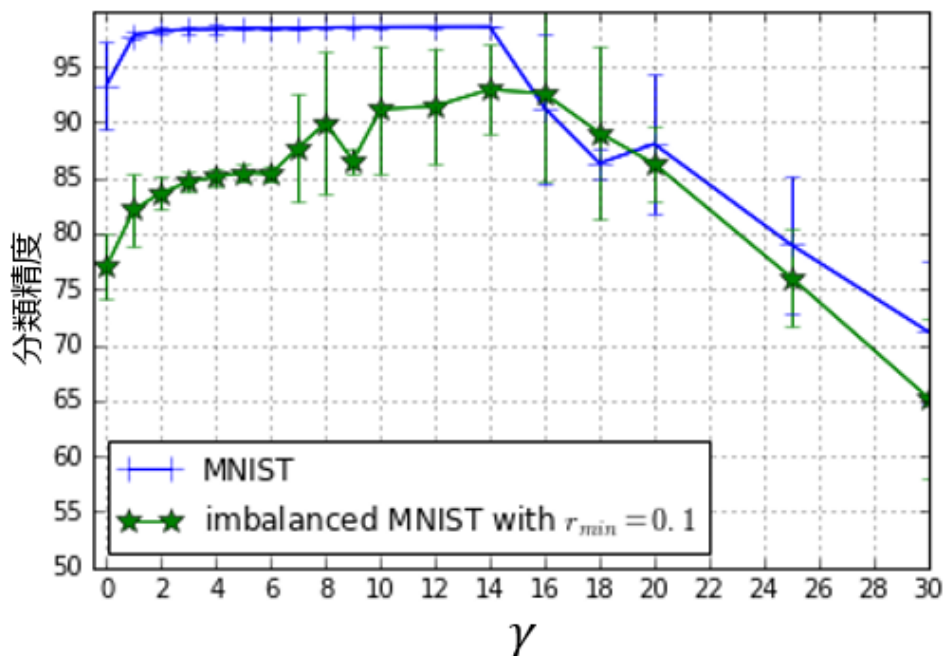


図10 ロス関数のパラメータ $\gamma$ の依存性

図10では $\gamma$ が2-6の値で安定して精度が高く、 $\gamma$ が大きくなるに従ってMNIST\_Lmbの標準偏差が大きく、バラつきが大きくなっている。 $\gamma$ が0の場合も大きすぎる場合にも精度は低下することから、RDECでは2つの項がバランスよく作用することで高い性能が得られることが確認できる。

図9で定性的に示したRDECの効果は、実験的にも確認することができる。図11は、MNISTのラベル0と6のデータを用いて、図9を実験的に再現した結果である。ここでは、少量の0（青）と大量の6（緑）を2つのクラスターに分類している。0のデータと6のデータの比は1:10となっている。ここでは、ネットワークの隠れ層の次元数を2に設定することで、潜在空間のデータ分布を可視化している。1点が手書き数字1枚を表し、青が0、緑が6を表している。三角は初期のクラスター中心を表している。

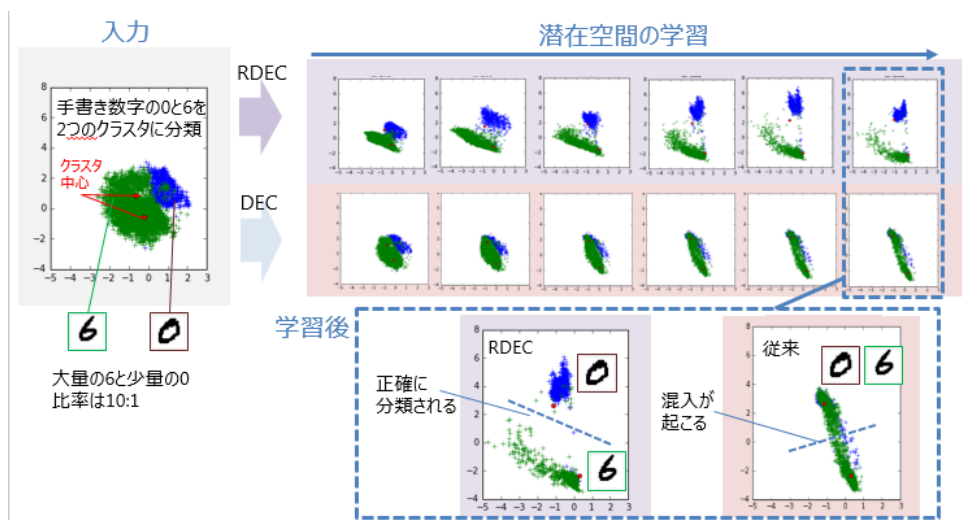


図11 RDECのロス関数の効果の実験的検証

図11では、DECではデータがクラスタ中心に凝縮され、同じ大きさのクラスタに分割されるのに対し、RDECでは近傍のデータが固まって動くため、正確に分類ができています。0と6の分類精度はDECの61.57%からRDECの99.37%に向上しており、RDECのロス関数の効果が実験的にも確認できます。

半導体製造の不良マップのように、現実のデータではしばしば少数のグループが重要な情報を持つ。RDECは、現実の課題解決と最新の深層学習技術が結びついた例であると考えられる。

---

## 5. 教師あり深層学習による既知不良マップ分類

---

本章では、深層学習手法を用いた既知不良の再発監視について、詳細を紹介する。第3章で述べたように、機械学習による既知の不良マップの分類は標準的な教師あり学習のアプローチであると言える。技術者が不良マップにラベルを付与することで、分類モデルを学習し、日々生産される製品を自動で分類することが可能になる。

深層学習の有効活用にはネットワークの構造やパラメータの設定が課題になる。深層学習は画像認識を初めとしたさまざまな分野で成功を収めているが、高い精度を達成するためにはパラメータを適切に設定する必要がある。製造現場への導入は必ずしも容易ではない。本章では、半導体工場で収集されたデータを用いた精度実験により、不良マップ分類に適した深層学習について考察する。

### 5.1 データセット

本稿では、半導体工場で収集された2カ月分のデータを用いて、注目する不良マップの自動分類精度を評価する。現実のデータではさまざまな不良マップが発生するが、本実験では図7の注目不良に示される、ウェハの4方向に放射上に不良が出現する放射上パターンを自動分類の対象とする。深層学習は放射状パターンとそれ以外のパターンの2値分類を行う分類モデルを学習する。現実の用途を模擬するために、2カ月分のデータを前半と後半に分割し、前半を訓練データ、後半を評価データとして用いる。

### 5.2 パラメータ設定

#### 5.2.1 ネットワーク構造

近年は、数多くの複雑なニューラルネットワークが提案されている。数年前であれば計算量が大きく製造分野への適用が困難であったネットワークも、計算機の発展と学習のノウハウの蓄積により、現実的な時間とコストで適用が可能になっている。

本稿では、比較的シンプルなニューラルネットを用いた実験を行う。ここでは5層から9層のConvolutional Neural Network (CNN) を用いて不良マップの分類モデルを学習する。不良マップは製品チップの数の画素を持つ画像と見なすことができるため、画像分野で大きな成果を上げているCNNが有効であると考えられる。また、不良の発生にはランダム性があり、同じ原因であってもウェハ上の不良の発生位置は揺らぐことから、CNNの使用が適切であるといえる。

ネットワーク構造は、入力層、畳み込み層、プーリング層、全結合層、出力層からなる5層のネットワーク構造を用いる。また、5層のネットワーク構造に畳み込み層とプーリング層のセットを追加して、7層、9層のネットワーク構造を用いる。活性化関数には、ReLU (Rectangle Linear Unit) を用いる。

表2は、ネットワーク構造と分類精度を表している。ここで、適合率はモデルが予測した放射状不良マップが実際に真であった割合、再現率はすべての放射状不良マップに対して真と予測された割合、F値は適合率と再現率の調和平均である。ベースラインとして、Support Vector Machine (SVM) を用いた分類精度を示す。

表2 CNNの放射状不良マップ分類精度

ネットワーク	F 値	適合率	再現率
SVM(Baseline)	0.884	0.889	0.880
5層	<b>0.920</b>	<b>0.934</b>	<b>0.905</b>
7層	0.906	0.930	0.882
9層	0.915	0.931	0.900

表2では、CNNは有力な機械学習手法であるSVMを上回る精度を達成している。ネットワークが深いほうが複雑な分類モデルを学習できるため、分類精度が向上することが期待されるが、表2では5層のネットワークが7層、9層のネットワークの性能を上回っている。これは不良マップの解像度やパターンが一般画像に比べるとシンプルであり、深い構造の分類モデルが不良マップに対して複雑すぎる可能性を示している。

深層学習は不良マップ分類に対しても有効であるが、シンプルなネットワーク構造が適していると考えられる。半導体製造においても、たとえば電子顕微鏡で撮像された画像の自動分類では、複雑なネットワーク構造で高い精度が得られることが報告されている ([22]等)。本稿の結果は不用意に深いネットワーク構造を用いるのではなく、問題に適したネットワークを選択することが重要であることを示唆している。

### 5.2.2 学習率の設定

深層学習は精度に影響する多くのパラメータが存在し、学習率は学習に最も影響の大きいパラメータの1つである。学習率は、1回の学習でモデルのパラメータをどの程度更新するかを決定するため、学習の精度と収束の速さに大きく影響する。本稿では、定数を用いた基本的な学習率から、近年では標準的になっている動的な学習率まで複数の学習率を比較し、マップ分類に適切な学習率について考察を行う。

ニューラルネットワークの学習は、ネットワークの重みを学習データに合わせて最適化するプロセスである。そのため、関数の勾配によりパラメータを更新して最適化を行う勾配降下法が学習に広く用いられる。

深層学習では確率的勾配降下法 (Stochastic Gradient Descent, SGD) と呼ばれる手法が用いられることが多い。SGDではデータをミニバッチと呼ばれる小グループ分割して、ミニバッチごとにパラメータを更新する。この方法により、学習を効率化するとともに、ミニバッチを分割にランダム性を持たせることで学習局所最適に陥ることを避ける効果があると考えられている。

SGDでは1回の学習でネットワークを更新する大きさを表す学習定数がハイパーパラメータになっている。学習定数は実験的に分類精度を確認しながら決定されることが多い。本稿では、学習定数を0.1, 0.01, 0.00 (SGD-0.1, SGD-0.01, SGD-0.001) としてそれぞれ分類精度を確認する。

本稿では定数による学習率の決定に加えて、動的な学習率についても評価を行う。学習過程で学習率を自動で調整するアルゴリズムとして、AdaGrad[18], RMSProp[19], AdaDelta[20], Adam[21]に注目する。

AdaGradとRMSPropは学習が進むにつれて、初期の学習率を減少させていく手法である。AdaGradは過去に大きく更新されたパラメータの更新率を小さくすることで学習率を更新する。RMSPropは、最新のパラメータの更新を優先して学習率を決定することで、AdaGradにおいて学習率が急速に0に減少してしまうという欠点に対処している。

AdaDeltaとAdamは学習過程で学習率を減少させるだけでなく、パラメータの更新に応じて増加させる手法である。AdaDeltaはAdaGradと同様に学習率を減少させるが、勾配の2次微分の値により学習率の値を増加させるアルゴリズムになっている。AdaDeltaにおいては、学習率の初期値が自動で決定されるため、設定が不要である。Adamは過去の勾配の履歴の1次モーメントと2次モーメントを元にパラメータを更新する。

図12は、各学習率の学習回数と誤分類率の関係を表している。誤分類率は評価データに対して放射状不良マップの分類の誤りの割合を表す。最高精度ではSGD-0.1が高い値を示しているが、収束の速さや汎用性を考えると、AdaGrad, AdaDelta, Adamも実用に向いていると考えられる。AdaDeltaは初期値を決定する必要がないという実用上の利点もある。

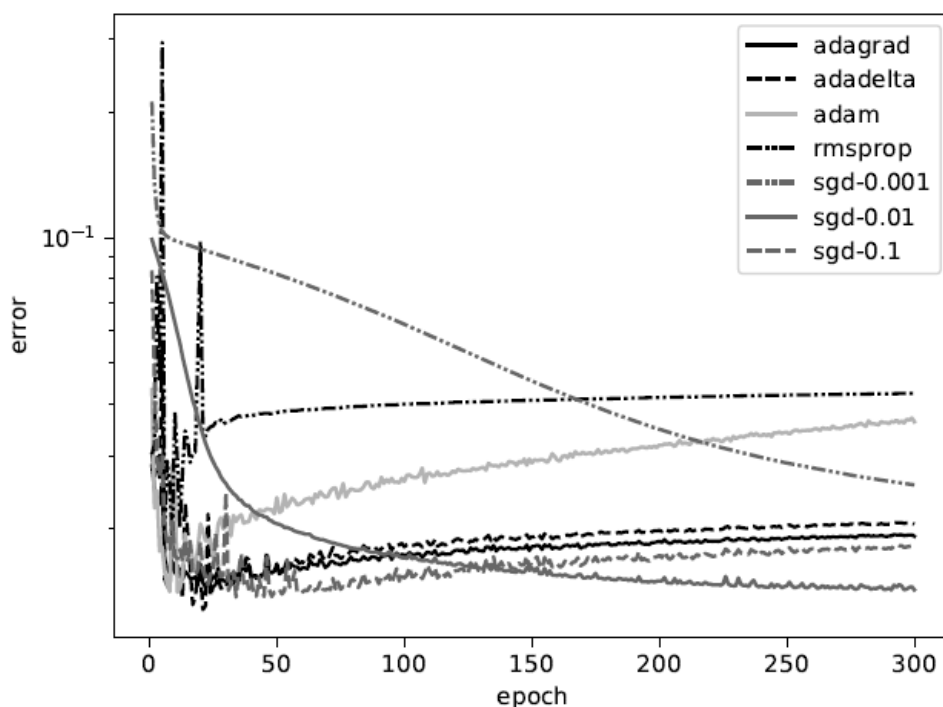


図12 各学習率の学習曲線 (評価データセットの性能)

機械学習では、一般的に学習回数が進むにつれて誤分類率は低下していくことが期待されるが、ここでは過学習による性能悪化が確認されている。AdaGrad, RMSProp, AdaDelta, Adamはいずれも学習が速く進み、100回未満の学習で過学習による精度の悪化が始まっている。

不良マップの分類では不良の発生頻度が少ないため、過学習が問題になる。不良マップは必然的にデータ量が少なくなり、マップのバリエーションも限られる。そのため、学習データに含まれる不良マップに特化した学習が行われると、評価データに含まれる若干パターンが変化した不良マップに対して分類精度が低下する。半導体工場では、製造装置状態の変化などで、不良マップのパターンが時期によって緩やかに変化することがあるため、過学習による精度劣化には十分注意する必要がある[10]。

---

## 6. 製造現場に展開した際の分類モデル学習手順

---

深層学習による不良マップの自動分類を製造現場に展開するには、ユーザである半導体技術者が簡単に分類モデルを学習する作業フローが必要である。

一般的に、ラベルを付与する作業には時間と労力がかかる。最新の半導体工場では日々大量の製品が生産されており、不良マップも多様で、時期によって発生状況が異なる。そのため、半導体の技術者がすべての不良マップを目視で確認して、注目不良マップか否かを判断してラベルを付与することは困難である。

一方で、ユーザの歩留解析の作業フローを考慮すると、注目した不良マップのみにラベルを付与することは現実的である。たとえば、技術者は歩留新聞（図2）で発見した少量の不良マップにラベルを付与することで、短時間で不良マップのリストを作成することができる。

歩留新聞の既知不良マップ分類では、不良マップが発生したウェハのリストのみから、学習を行う仕組みを導入している。図13は、歩留新聞の既知不良マップ分類を行う際のユーザの作業フローの例を表している。図13では、ユーザが注目する不良マップのリストをアップロードすると、システムがウェハの検査時期などのメタ情報を用いてその他の不良マップを自動で収集して学習を行う。通常の教師あり学習では、注目する不良マップ（正例）と、それ以外のマップ（負例）のすべてにラベルを付与する必要があるが、ここでは、注目する不良マップ以外のマップを自動的に選択することで、負例のラベルを付与する必要がないフローが導入されている。

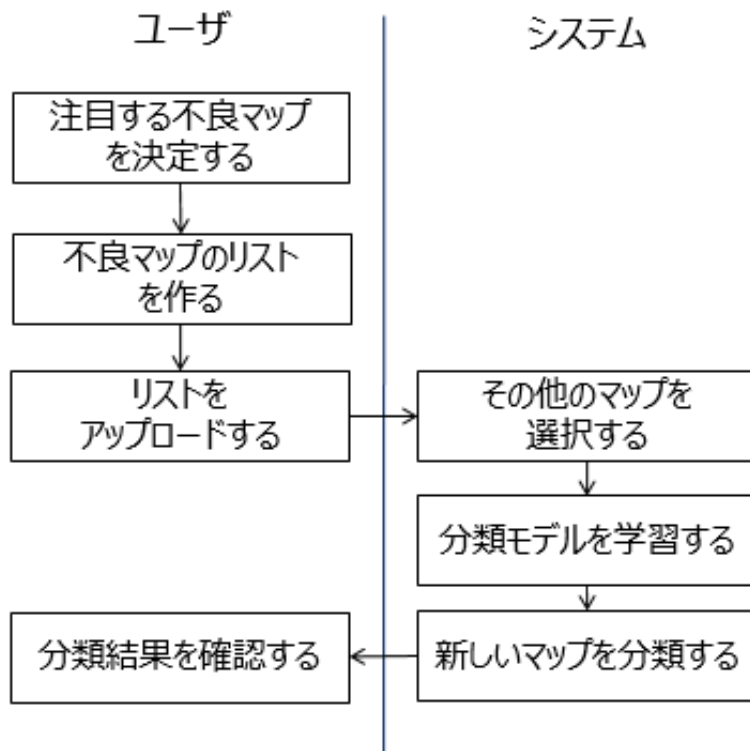


図13 既知不良マップの分類モデル学習フロー

第5章において用いた注目不良マップの学習データに対する学習時間は、GPUを用いた場合に約4分となっており、ユーザがリストを登録すれば、すぐに監視を開始することができる。この作業フローにより、機械学習に不慣れなユーザであっても、大きな負荷なしに図3で示したインタフェースによる監視を行うことができる。

## 7. まとめと今後

歩留新聞により、毎朝、前日までの全製品・全不良カテゴリの不良マップ、発生トレンド、原因候補の情報を一覧で確認することを可能にした。不良マップと原因装置候補を技術者に提示することで、不良1件あたりの解析時間を平均6時間から2時間に短縮した。

歩留新聞を実現する深層学習技術として、深層クラスタリング技術RDECと、深層学習CNNのチューニングについて紹介した。機械学習の実応用には、古典的技術から最新技術までを俯瞰した上で、問題に適した技術の選択が重要である。我々は、生産性向上の課題を明確にし、適材適所に機械学習・データマイニング技術を適用して課題を解決しながら、最先端の技術開発を行っていく。

### 参考文献

- 1) 中田康太, 折原良平: ビッグデータを活用した歩留まり解析支援システム“歩留まり新聞”, [https://www.toshiba.co.jp/tech/review/2018/03/73\\_03pdf/a05.pdf](https://www.toshiba.co.jp/tech/review/2018/03/73_03pdf/a05.pdf)
- 2) Nicolao, G. D., Pasquinetti, E., Miraglia, G. and Piccinini, F. : Unsupervised Spatial Pattern Classification of Electrical Failures in Semiconductor Manufacturing, in Proc. Artificial Neural Networks Pattern Recognition Workshop, 125-131 (2003).
- 3) Chien, C. F., Wang, W. C. and Cheng, J. C. : Data Mining for Yield Enhancement in Semiconductor Manufacturing and an Empirical Study, Expert Systems with Applications, vol33, 1, pp.192-198 (2007).



- 4) MacQueen, J. : Some Methods for Classification and Analysis of Multivariate Observations, in Proceedings of The Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol1, 14, pp.281-297 (1967).
- 5) Borgelt, C. : An Implementation of The FP-growth Algorithm, in Proceedings of The 1st International Workshop on Open Source Data Mining : Frequent Pattern Mining Implementations, 1-5 (2005).
- 6) Zhou, L., Zhong, Z., Chang, J., Li, J., Huang, J. Z. and Feng, S. : Balanced Parallel Fp-growth with Mapreduce, Information Computing and Telecommunications.
- 7) Liao, C. S., Hsieh, T. J., Huang, Y. S. and Chien, C. F. : Similarity Searching for Defective Wafer Bin Maps in Semiconductor Manufacturing, IEEE T. Automation Science and Engineering, vol 11, 3, pp.953-960 (2014).
- 8) Adly, F., Yoo, P. D., Muhaidat, S., Al-Hammadi, Y., Lee, U. and Ismail, M. : Randomized General Regression Network for Identification of Defect Patterns in Semiconductor Wafer Maps, IEEE Transactions on Semiconductor Manufacturing, vol 28, 2, pp.145-152 (2015).
- 9) Wu, M. J., Jang, J. S. R. and Chen, J. L. : Wafer Map Failure Pattern Recognition and Similarity Ranking for Large-Scale Data Sets, IEEE Transactions on Semiconductor Manufacturing, vol 28, 1, pp.1-12 (2015).
- 10) Nakata, K., Mizuoka, Y., Takagi, K. and Orihara, R. : A Comprehensive Big-Data-Based Monitoring System for Yield Enhancement in Semiconductor Manufacturing, IEEE Transactions on Semiconductor Manufacturing 30.4, 339-344 (2017).
- 11) Tao, Y., Takagi, K. and Nakata, K. : RDEC : Integrating Regularization into Deep Embedded Clustering for Imbalanced Datasets, Asian Conference on Machine Learning (2018).
- 12) Hinton, G. and Salakhutdinov, R. : Reducing The Dimensionality of Data with Neural Networks. Science, 313 (5786), 504-507.
- 13) Xie, J., Girshick, R. and Farhadi, A. : Unsupervised Deep Embedding for Clustering Analysis, In International Conference on Machine Learning, 478-487.
- 14) Yang, B., Fu, X., Sidiropoulos, N. D. and Hong, M. : Towards K-means-Friendly Spaces : Simultaneous Deep Learning and Clustering, ArXiv Preprint ArXiv : 1610.04794 (2016).
- 15) Jiang, Z., Zheng, Y., Tan, H., Tang, B. and Zhou, H. : Variational Deep Embedding : An Unsupervised and Generative Approach to Clustering, ArXiv preprint, ArXiv : 1611.05148 (2016).
- 16) Hu, W., Miyato, T., Tokui, S., Matsumoto, E. and Sugiyama, M. : Learning Discrete Representations Via Information Maximizing Self Augmented Training, ArXiv, preprint ArXiv : 1702.08720 (2017).
- 17) Miyato, T., Maeda, S., Koyama, M. and Ishii, S. : Virtual Adversarial Training : A Regularization Method for Supervised and Semi-supervised Learning, ArXiv preprint ArXiv : 1704.03976 (2017).
- 18) Duchi, J., Elad, H. and Singer, Y. : Adaptive Subgradient Methods for Online Learning and Stochastic Optimization, Journal of Machine Learning Research 12. Jul, 2121-2159 (2011).
- 19) Tieleman, T. and Hinton, G. : Lecture 6.5-rmsprop : Divide The Gradient by a Running Average of Its Recent Magnitude, University of Toronto, COURSERA, Neural Networks for Machine Learning, vol 4, 2, 26-31 (2012).
- 20) Zeiler, M. D. : ADADELTA : An Adaptive Learning Rate Method, ArXiv Preprint

ArXiv : 1212.5701 (2012).

21) Kingma, D. and Ba, J. : Adam : A Method for Stochastic Optimization, ArXiv Preprint ArXiv : 1412.6980 (2014).

22) Imoto, K. et al. : A CNN-based Transfer Learning Method for Defect Classification in Semiconductor Manufacturing, International Symposium on Semiconductor Manufacturing (2018).

**中田 康太** (非会員) kouta.nakata@toshiba.co.jp

(株) 東芝 研究開発本部 研究開発センター アナリティクスAIラボラトリー 主任研究員. 2006年に東京大学 理学系研究科 地球惑星科学専攻から (株) 東芝に入社. データマイニング, 機械学習技術の研究開発に従事. 博士 (理学). 人工知能学会会員.

**高木 健太郎** (非会員) kentaro1.takagi@toshiba.co.jp

(株) 東芝 研究開発本部 研究開発センター アナリティクスAIラボラトリー. 2015年に京都大学 理学研究科 物理学・宇宙物理学選考から (株) 東芝に入社. 機械学習技術の研究開発に従事. 博士 (理学).

**陶 亜玲** (正会員) yaling1.tao@toshiba.co.jp

(株) 東芝 研究開発本部 研究開発センター アナリティクスAIラボラトリー. 2016年に筑波大学 システム情報工学研究科 社会工学専攻から (株) 東芝に入社. 機械学習技術の研究開発に従事. 博士 (工学).

採録決定 : 2019年1月24日

編集担当 : 石黒 剛大 (三菱電機 情報技術総合研究所)