**機械学習による**

**フォトマスクの寸法分布予測と**

**歩留まり改善**

(M技研)(P技C)[P二開](マ技) 宇賀神　邦裕

(M技研)(P技C)[P二開](マ技) 中　真人

目次

1. 背景　……………………………………………………………………3
   1. フォトマスクの製造ライン
   2. フォトマスクの不良原因
   3. フォトマスク製造における寸法分布予測の課題
2. フォトマスク製造への機械学習の導入　…………………………………7

２-１機械学習の問題設定

２-２機械学習の種類

1. 機械学習の実験条件　……………………………………………………10

３-１特徴量エンジニアリング

３-２実験条件

1. 1st Trialと考察　－失敗例－…………………………12
2. 2nd Trialと考察　－改善例－……………………………14
3. 更なる高精度化に向けた検討 　……………………X
4. 今後の展望 ………………………………………………………………X

謝辞　…………………………………………………………………………X

参考文献　………………………………………………………………X

Appendix　………………………………………………………………X

**1.　背景**

近年、AI（**A**rtificial **I**ntelligence）が様々な分野で活用されるようになっている。AIとは、人間と同様の知識を実現させようという取り組みやその技術のことを指す。AIという概念は従来からあったが、IoT(**I**nternet **o**f **T**hings)によるビッグデータの入手や、半導体メモリ容量の増加、高速演算が可能なハードウェア能力の向上などによって、近年実現可能な技術となり急速に開発や運用が進められている。このような流れの中で、製造業においてもAIの活用が盛んになっている。製造業におけるAIの導入は、少子高齢化に伴う人材不足、オペレーターの人件費削減、歩留まり向上などの生産管理の高精度化といったニーズを満たす。そのため、弊社のウェハファブでも近年開発が急速にされており、省人化や歩留まり向上への寄与がなされている。しかしながら、我々が開発に携わっているフォトマスクの生産ラインでは、AIの十分な活用ができていないのが現状となっている。

## 1-1　フォトマスクの製造ライン

　フォトマスクは半導体回路の原版に相当し、透明なガラス基板の上に光を通さない薄膜でパターンが形成されたものである。パターンをウェハに転写するためには、リソグラフィ技術[[1]](#footnote-1)が使われる。リソグラフィでは再現性よくパターンが転写するため、フォトマスクのパターンは正確に作成する必要がある。フォトマスクの製造フローを図１に示す。

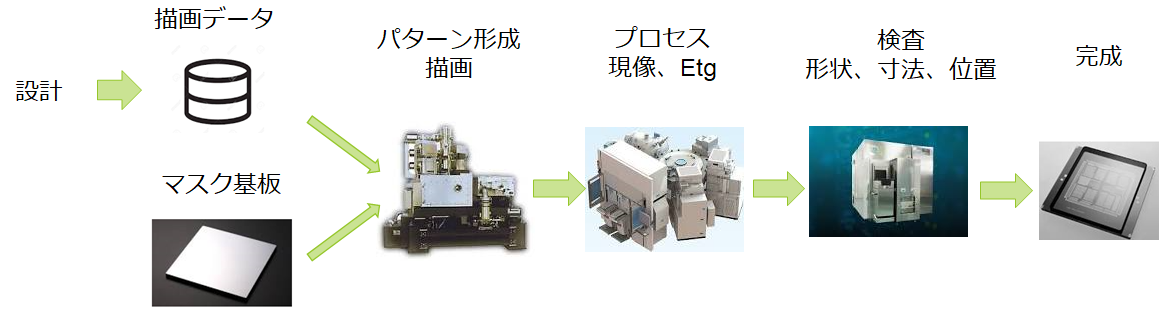


図１．フォトマスク製造フロー

まず設計データをもとに描画データを用意し、マスク基板にパターンを描画する。次にプロセス処理してパターンを形成する。最後に検査を行い、良品のフォトマスクが完成品として出荷される。原版であるフォトマスクの製造フローは品質を保証した生産体制になっていることが特徴である。品質の悪いフォトマスクが出荷されると、ウェハの歩留まりを落とすことになるため、フォトマスクは要求される品質を満たしつつ、フォトマスク自体の歩留まりも向上させる必要がある。

## 1-2　フォトマスクの不良原因

フォトマスクに要求される品質は、半導体メモリの高集積化に伴い、年々厳しくなっている。半導体メモリの集積化のベクトルは従来の2次元平面内での微細化の方向から、3次元空間での積層方向へと向きを変えつつある。そのため、ウェハパターンの寸法精度や各レイアウト間の合わせ精度は高精度が要求され、これ伴いフォトマスクに要求される品質（精度スペック）も高くなっている。

我々が開発に携わっているフォトマスクはDTF（**D**ainippon **T**oshiba **F**ine Electronics）の川崎の第一製造課と北上の第二製造課にて製造されている。各世代のフォトマスクにはリソグラフィ工程における尤度(リソマージン)から計算された精度スペックが定められている。2010年から2020年にかけて川崎にて製造したフォトマスクのうちスペック未達となった不良品の原因の推移を図２に示す。

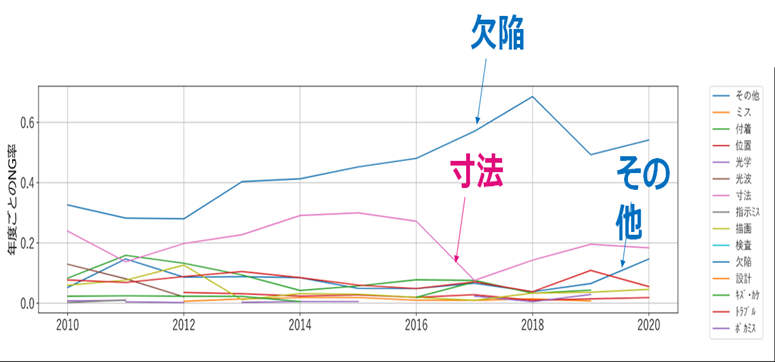


図２．フォトマスク製品の不良原因

図２から、不良品の原因は年代であまり変化がなく、欠陥が約50%、寸法が約20％の発生率を占めていることが確認される。そのため、フォトマスクの歩留まり改善に効果的な対策は、欠陥対策と寸法対策であると考えられる。本報告では寸法対策に焦点を当てることとした。

## 1-3　フォトマスク製造における寸法分布予測の課題

本検討を進めていくにあたり、まず、フォトマスクの寸法精度の歩留まりが良好でない世代の調査を行った。その結果、北上の製造ラインのZ規格KrF露光マスクの寸法歩留まりが良くないことが確認された。該当の寸法精度のトレンドデータ（2021年５月～2022年3月）を図３に示す。

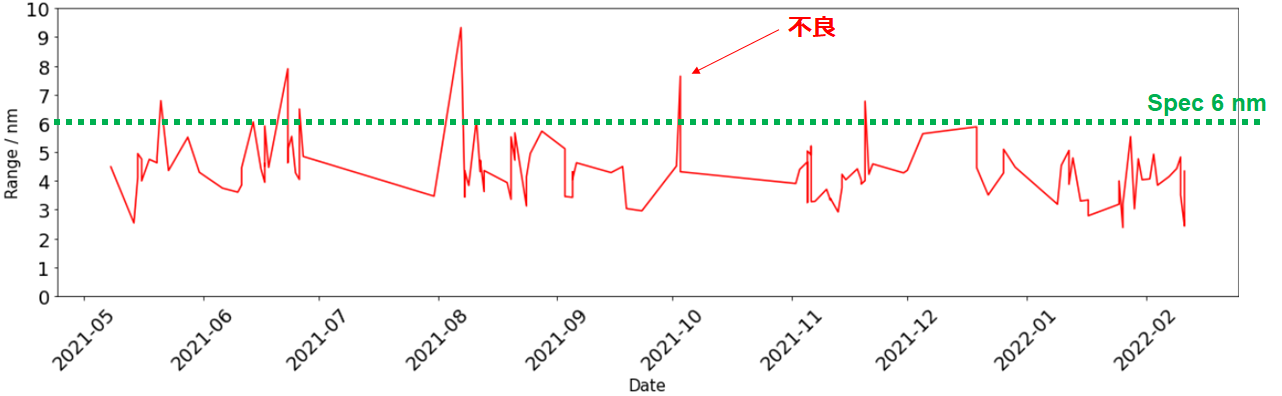


図３．北上Z規格KrF露光マスク寸法精度

Z規格とはフォトマスクの規格であり、32nmノード相当の世代にあたる。この世代のマスク上のパターン寸法精度はRange(Max-Min)：6nmと定められており、図３のトレンドデータからもスペックを越えた不良品が散発していることが確認された。工程能力を以下の式(1)から計算すると、Cpk=0.59（不良率：7.6%）となる。

Cpk＝（上限規格 – 平均値）／（３×標準偏差） (1)

このような不良率の高さは、フォトマスク製造においてパターンレイアウトを描画する描画装置の寸法補正の予測精度が低いことに起因すると考えられる。フォトマスクの寸法精度コントロール手順を図４に示す。

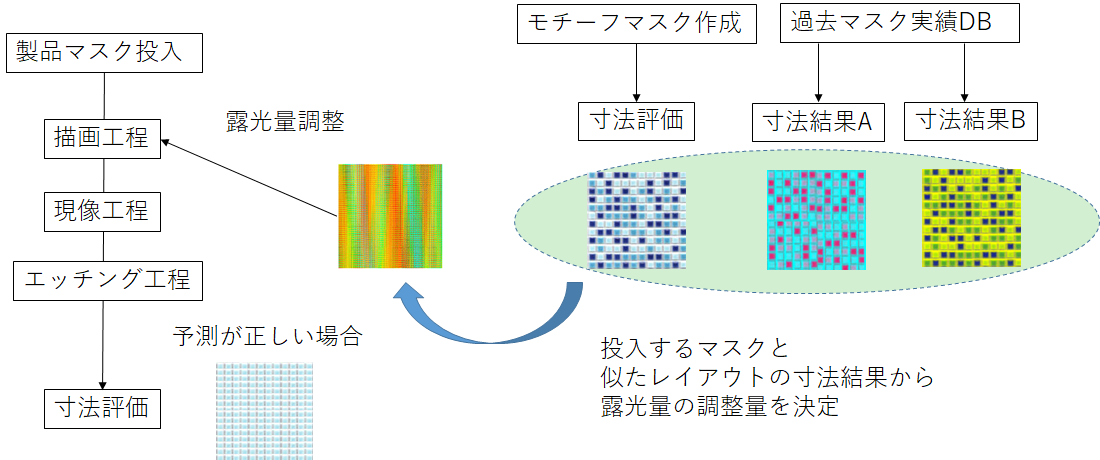


図４．フォトマスクの寸法精度コントロール手順

図４左はフォトマスクの寸法制御に関連する製造フローを示している。フォトマスクの寸法は、主に、描画工程、現像工程、エッチング工程により決定づけられるものであるが、その中でも、描画装置は露光量で寸法調整が可能な装置である。

現状の運用の手順について説明する。新たにフォトマスクを作成する際には、まず、その世代を代表するモチーフマスクを作成して寸法評価をしておくか、過去の製品結果から類似のフォトマスクを検索してその寸法分布を入手する。得られた寸法分布を補正するように露光量を調整するレシピを描画装置に設定する。事前に予測した寸法分布が合っている場合には、フォトマスクの寸法は正しく補正されて面内の寸法分布は均一になる。

この寸法分布を予測する手順には２つの問題点がある。一つは類似したフォトマスクを検索するところである。フォトマスクは少量多品種であるため、同一のマスクレイアウトが再生産されることはほとんどなく、類似したフォトマスクを探すことは困難であるということ。もう一つは時間変動の影響である。予測する寸法結果は過去の実績から参照するため、各製造工程の装置状態の変動（多くの場合時間的な変動であらわれる）を加味できていないことである。

　これらの問題点を解決するための手法として、我々は機械学習を活用することを検討した。機械学習は過去のデータを学習することで、過去の経験に基づいた未来予測を可能にする技術である。一つ目の問題点の類似のフォトマスクを検索することに関しては、学習したデータを利用して未知のレイアウトの寸法分布を予測することが可能である。また、二つ目の時間軸変動の影響に関しては、学習データにフォトマスク作成の時間情報を取り込むなどの工夫により表現できると考えられる。

そこで、本報告では、北上Z規格KrF向けフォトマスクの寸法精度の歩留まり改善のため、機械学習を用いてフォトマスクの寸法分布を予測した検討結果を報告する。予測した寸法分布を補正するように露光量調整を行うことでフォトマスクの寸法分布の不良率を7.6%から1%以下にすることを目標とした。

第二章では機械学習について説明し、第三章では機械学習の予測精度を構築するために重要な特徴量の選定と、機械学習の計算機環境について述べる。第四章では、XXXX。

**２.**フォトマスク製造への機械学習の導入

## 2-1機械学習の問題設定

機械学習はデータの与え方から教師あり学習と、教師なし学習とに分けることができる。

教師あり学習は、入力データと対になる正解の出力データとを与えて学習する手法である。例えば事前に大量の動物の画像を用意し「これはネコ」、「これはイヌ」といったラベルをつけて学習をおこなう。学習済みのモデルでは、未学習の画像を入力したときに、その画像がネコかイヌかを判定できるようになる。

教師なし学習は、入力データに対する正解ラベルがないデータの学習方法である。入力データ間の類似度を分析しグルーピングを進めることが学習の方針になる。

本報告では過去のフォトマスク製品の寸法値を教師データとして学習し、新たに作成するフォトマスクの寸法分布を予測するため、教師あり学習として扱う。また、上記では「ネコ」と「イヌ」といった分類問題を例としたが、フォトマスクの寸法値は計測した数値である。したがって数値を予測する回帰問題として扱う。

## 2-2機械学習の種類

機械学習における学習の進め方には様々なアルゴリズムがある。本報告では教師ありデータの回帰問題のアルゴリズムとしてよく使われるニューラルネットワークと決定木について比較検討をおこなう。それぞれの学習のアルゴリズムについて以下に説明する

i）ニューラルネットワーク

ニューラルネットワークとは人間の脳の中にある神経細胞（ニューロン）とその繋がりを数式によるモデルで表現したものである。図５に神経細胞の信号伝達と人工ニューロンのモデルを示す。

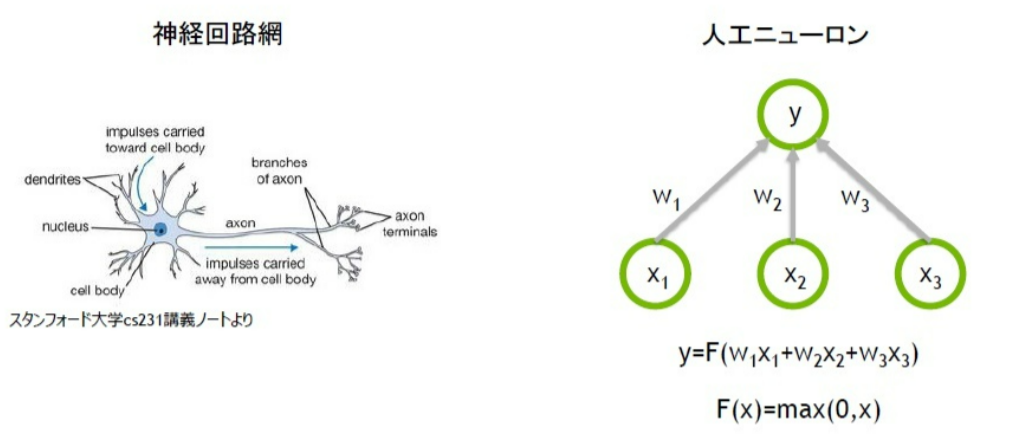


図５．神経回路網と人工ニューロン1)

図５左の神経細胞は樹状突起を経由して他の多くの神経細胞からの信号を受け取る。それぞれの信号は神経伝達物質の量によって伝わりやすさが異なっている。神経細胞が受けた電気信号の総電気量が一定値を超えると次の神経細胞へ信号を伝える仕組みになっている。

一方で、右の人工ニューロンでは入力x1～x3に対して重みづけの係数w1~w3を用意する。次にそれぞれを掛け合わせた線形和を計算し、最後に活性化関数で総和値の変換をおこなう。最後の活性化関数を入れることで非線形の表現が可能となり複雑なモデルを表現できるようになる。

ニューラルネットワークにおいて学習を進めていく方法は、予想された出力yと教師データとの誤差を最小にするように重みづけの係数w1~w3を調整することである。詳細の計算方法はappendixに記載する。

ii）決定木

　決定木とはデータを木構造で分類していくことで、そのデータの構造を学習する機械学習方法である。データを木構造で分割して正解に近くなるように各分岐条件を設定する。例として図６に受信したメールがスパムメールかどうかを判定する決定木を示す。

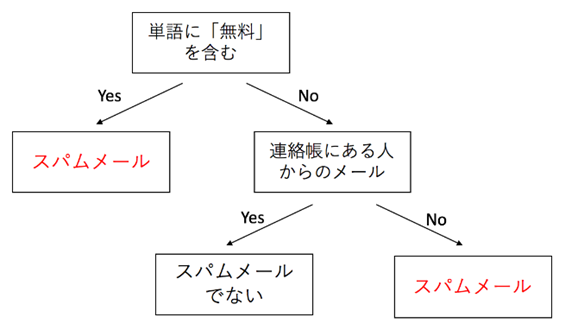


図６．スパムメールを判断する決定木

図６では『単語に「無料」を含むか』と『連絡帳にある人からのメールか』の２種類の入力データが与えられたときに『スパムメールであるか』の判定をおこなっている。どちらの入力を先に判定するかは、分割したグループのデータ情報量が少なくなるほうが優先して選択される。また、複数の決定木が候補となる場合には、木構造がシンプルな形になるものが選択される。

決定木にはいくつかの構造の組み方があるが、ここではLightGBM（Light Gradient Boosting Machine）を例に説明する。LightGBMは勾配ブースティングとよばれる決定木のアルゴリズムの一つである。勾配（Gradient）とは微分して求められた傾きのことで、予想値と正解値の誤差の微分値に相当する。ブースティング（Boosting）の手順を図７に示す。

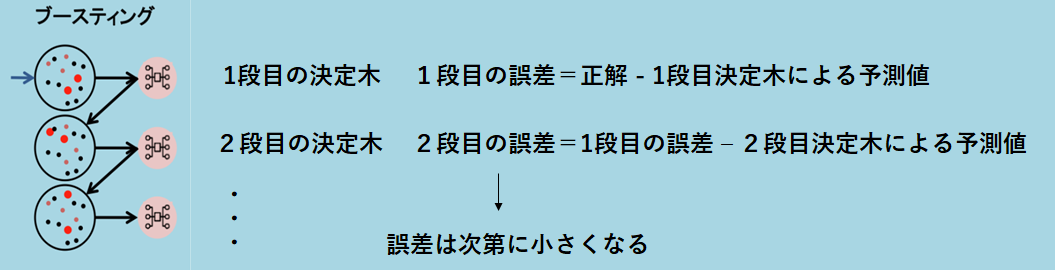


図７．ブースティング2)

最初の1段目の決定木モデルで正解値を予測するモデルを学習する。単体の決定木モデルではデータを完全に表現することは難しく、予想値と正解値との間には誤差が生じる。次の２段目の決定木モデルではこの1段目の誤差を学習する。この作業を繰り返しおこない、全決定木モデルの予測値の総和を最終の予測値とする方法がブースティングである。これにより、単体では精度の低いモデルでも複数組み合わせることで精度の高いモデルを作ることができる。

次に条件分岐の学習の進め方について図８を用いて説明する。

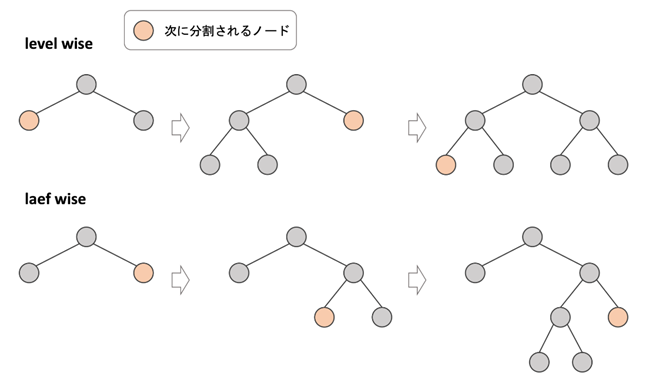


図８．分岐の進め方

一般的な決定木の分岐の進め方は図８の上段にあるlevel wiseと呼ばれる方法である。level wiseでは同じ階層まで分割を進めながら学習をおこなう。一方、LightGBMではleaf wiseとよばれる手法が使われている。各葉の中の誤差の大きいものから選択的に分岐をおこなっていく。これにより同じ計算量でも高精度な分岐が可能となる。今まで述べてきたように決定木モデルのなかでもLightGBMは予測精度が高く、計算速度も速いといった特徴がある。本報告でも決定木モデルのベースとしてLightGBMを使って検討を行う。

　本節で紹介したように学習のアルゴリズムによって機械学習のモデルの構成は異なる。ニューラルネットワークでは出力ｙを入力ｘの関数として表す意味合いが強い。そのため連続的に変化するｙを表現しやすいモデルとなっている。一方で決定木ではデータの全体の構造を木構造として表す。これによりデータ全体の持つ特性が表現されやすい。分岐の回数を増やすことで、局所的な変動なども表現することができる。どちらのモデルが優れているかについては一概には判別が難しいため、それぞれのモデルを検討しながら機械学習の精度を向上させることが必要である。

次の第3章では、機械学習の予測精度を構築するために重要な特徴量の選定と、機械学習の計算機環境について述べる。

**3.** 機械学習の実験条件

## 3-1　特徴量エンジニアリング

特徴量とは機械学習モデルをつかって出力を予測するときに用いる入力データのことである。どのような特徴量を使って機械学習を進めるかによって、モデルの予測精度や汎用性が決まってくる。精度の高い予測モデルを構築するためには、その予測値に影響の大きい特徴量を用意する必要がある。そのため、本検討を進めるにあたり機械学習で用いる効果的な特徴量を選定した。

本検討ではフォトマスクの面内の寸法分布を予測しようするため、面内の寸法測定点を示す位置情報が特徴量として効果的と考えられる。通常のフォトマスクの寸法測定箇所は測定時間の都合で面内200点程度に制限されているため、1枚の製品測定結果から寸法分布全体の傾向を判断するのは困難である。そのため、多数のフォトマスクの製品の寸法値を同じ座標平面にプロットすることで面内の寸法分布の傾向を調査することにした。図9は、2020年11月～2021年5月で製造した約200枚のフォトマスクの寸法値の偏差を同一平面上にバブル図で示したものである。青は基準に対して寸法が太る方向、白は基準に対して寸法が細る方向を表しており、バブルの大きさで偏差量を表している。

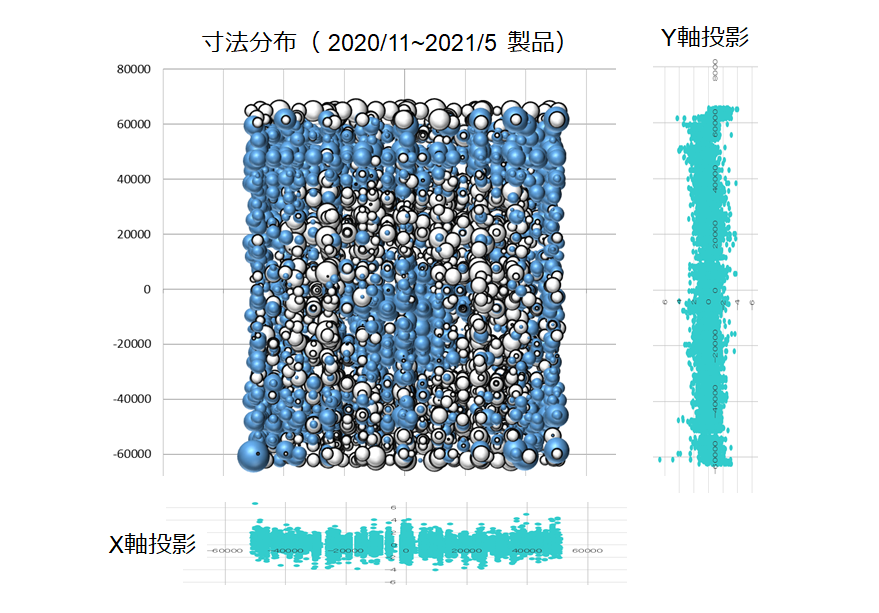


図9. フォトマスクの面内寸法分布

フォトマスクごとに寸法の計測個所は異なるが、異なるフォトマスクの寸法結果を重ねて表示することで、全体の寸法変動の傾向を把握することができる。図９から、①上下端部では寸法が急激に平均値を下回る様子と②同心円状にドーナツ型の寸法変動が確認できる。このようなマスク面内の寸法変化を表現するため、座標の高次項を特徴量に加えることで端部の大きな寸法変動を表現することを考えた。また、同心円の構造を表現するために、極座標の直交座標系であるゼルニケ級数を特徴量に加えることにした。

さらに調査を進める中で、フォトマスクのパターンの密度を表す被覆率と面内の寸法変動との間に関係性があることが分かった。図10に被覆率の異なる2種類のマスクの寸法分布を示す。

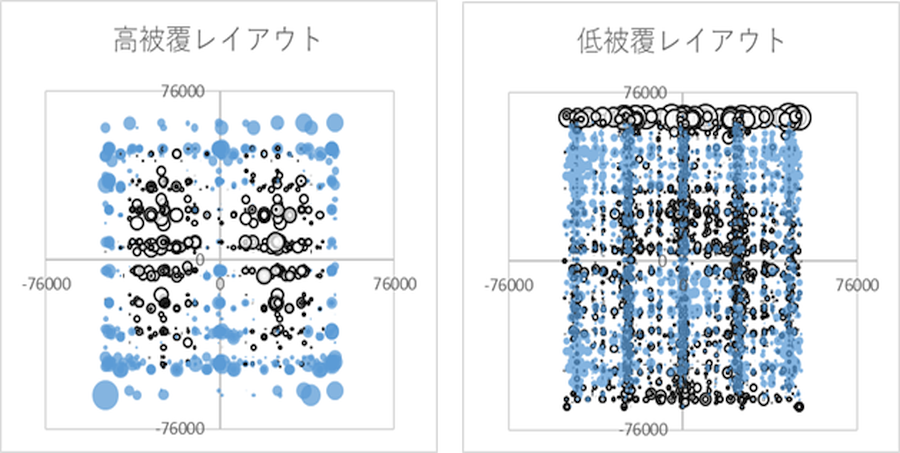


図10. 被覆率の違いとフォトマスクの寸法分布の関係

図10の寸法偏差の分布は図９と同様に複数のフォトマスクの製品結果を重ねた結果である。左側はパターン密度の高い高被覆率のフォトマスクであり、右側は低被覆率のフォトマスクの寸法分布である。高被覆率のレイアウトでは２つの縦長の楕円の分布が見えている。一方、低被覆率のレイアウトでは、上下端部では寸法が急激に平均値を下回る様子と、真ん中に小さなドーナツ状の分布になっていること確認できる。これはフォトマスクのパターン被覆率が増えると、①描画工程における露光時のかぶりや、②加工工程における被覆率依存のローディングにより、フォトマスクの寸法変動が大きくなることが知られている。1)、2)　このようなレイアウトの特徴も寸法分布に影響を与えるため、フォトマスクの被覆率、レイヤーの種類などの情報も特徴量に加えることとした。

特徴量には「座標」のように数字で表すことのできる定量的なデータのほかに、「レイヤーの種類」のような文字で表す定性的データも使うことができる。後者の定性的なデータはカテゴリーデータともよばれ、数値での処理ができないためラベルとして学習する。

このような検討を重ねた結果、最終的には38種類の特徴量をつかってモデルの学習をすすめることにした。使用した特徴量の一覧を図11に示す。

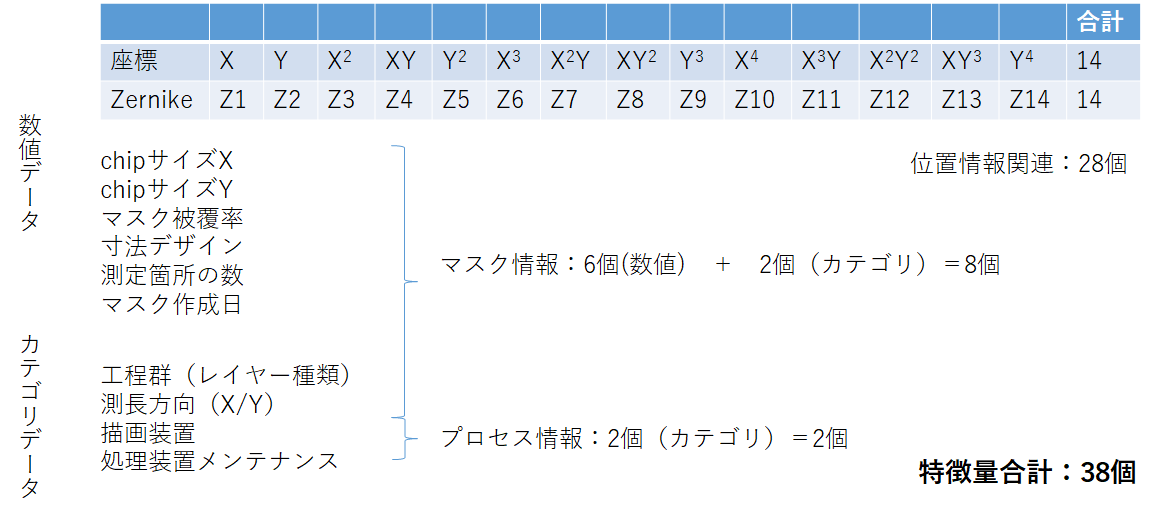


図11. 本検討で用いた特徴量の一覧

## 3-2　機械学習の計算機環境

機械学習の計算にはAWS(Amazon Web Services)を利用した。8CPUと1GPU（T4）を利用して高速に演算することが可能である。

機械学習のプログラミングにはPython(Python3.9.2)を用い、学習モデルの構築にはPythonのライブラリーを利用した。ニューラルネットワーク用にはPyTorchを用いた。LightGBMでは専用のライブラリーをインストールして使用した。図12に主に今回使用したライブラリーのバージョン番号を示す。

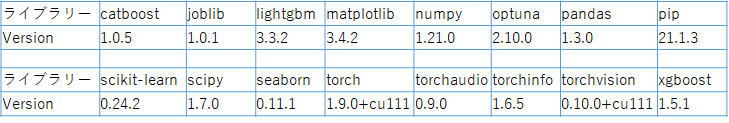


図12. 使用したライブラリーとバージョン番号

次の第4章では、第3章で選定した特徴量を用いた機械学習の結果と課題にについて述べる。

**4.**1st Trialの結果と考察　－失敗例－

## 4-1　モデル学習の手順

機械学習では与えられたデータの法則性を学習していくが、学習が過剰に進むと、過学習という問題が発生する。過学習とは学習データに対して極端な忠実性を持つ一方、未学習のデータに対して予測が外れている状態を指す。これを回避するために以下の方法が用いた。まずデータを学習用データ（Train Data）と検証用データ（Validation Data）とに分ける。次に学習用データでモデルの学習をおこない、検証用データでモデルの妥当性を評価する方法である。それぞれの学習セットはfoldとよばれ、それぞれのfoldでモデルを学習する。今回は全体の20%を検証用データとして図13のようにデータを分割してモデルを作成した。

例えば、fold3では全ての製品データを5つのグループに分け、左から3番目以外の４つのグループを学習用データとしてモデルの学習をおこない、左から3番目のグループを用いてモデルの妥当性を検証している。

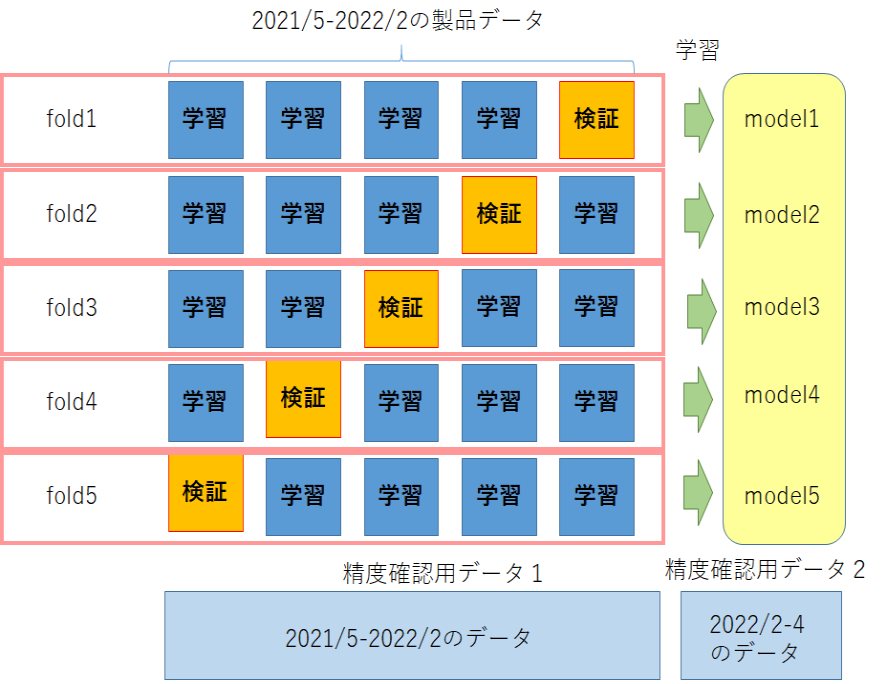


図13. 1st Trialモデルデータ分割方法

それぞれのfoldでは、検証データと最も合うように学習モデルが作成される。すべての学習を終えると5個のモデルが作成できるので、最終的な寸法予測値には5個のモデルで算出した寸法値の平均値を用いた。

フォトマスクの製品は１章で紹介したようにあらかじめ寸法分布を経験的に予測し、予測した寸法分布を補正している。今回の機械学習の見積もりはこの経験的に予測していた部分に相当するため、機械学習を適用した場合の製品寸法予測は、補正前の製品結果から予測した寸法値を差し引いたものを計算することになる。

## 4-2　失敗例と課題

このようにして機械学習を適用した場合の製品寸法予測（simulation）と、実際の製品結果（product）の比較を図14に示す。評価対象は図13の精度確認用データ１(学習に用いたデ―タ)になる。

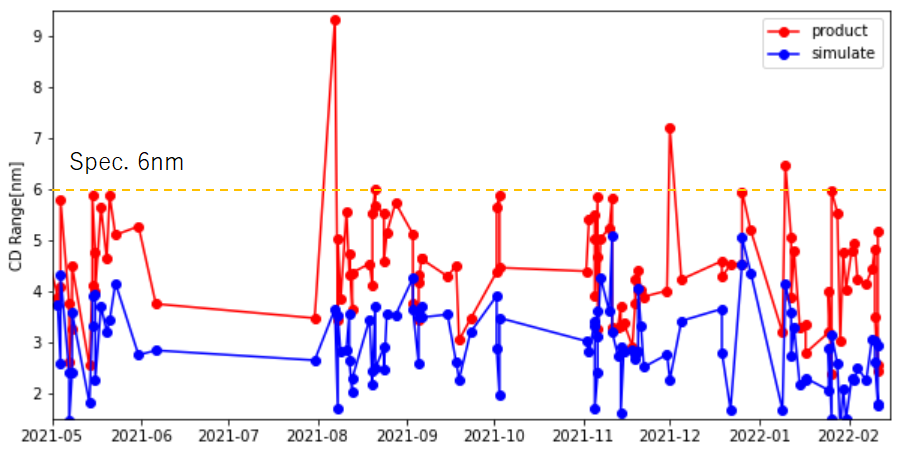


図14. 製品結果(product)と予測分布での補正後見積もり(simulation)

製品の実績（赤）に対して、予測した寸法値を補正した見積もり（青）の寸法Rangeが大幅に改善していることが確認された。CD Range値の中央値も4.4nmから2.8nmと改善している。このモデルを使って、学習データで使用していない2022年2月～2022年4月に作成されたフォトマスクの寸法精度を予測した結果を図1５に示す。評価対象は図13の精度評価用データ２（未知のデータ）になる。

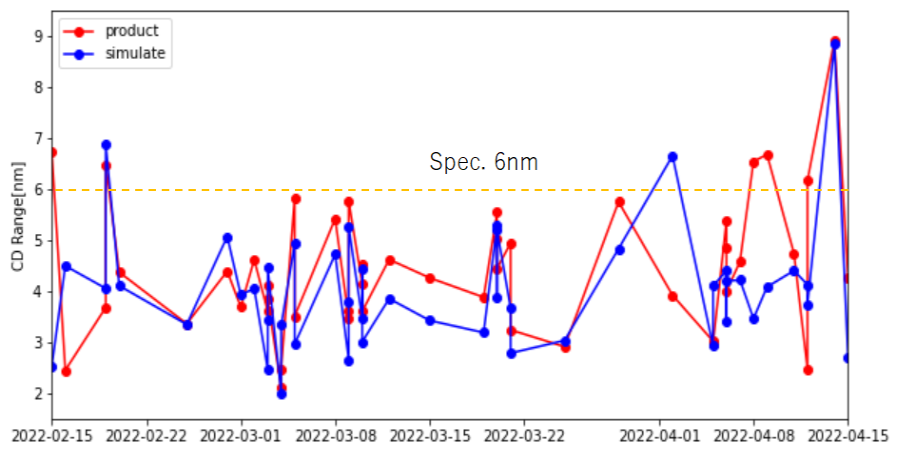


図1５. 学習データで使用していないフォトマスクの製品結果と補正後見積もり

モデル学習に用いた精度確認用データ１（202１/5-2022/2）に対しては寸法精度の改善が見られたのに対して、未知のデータにあたる精度確認用データ2(2022/2-2022/4)に対しては製品の実績（赤）と機械学習で予測した寸法値を補正した見積もり（青）はほとんど変わっていない。むしろ一部のフォトマスクでは現在の製造方法より、寸法分布が劣化していることが分かる。CD Range値の中央値も4.６nm（赤）から4.1nm（青）と改善が見られない結果となり、4-1で述べた過学習の状態（学習データに対して極端な忠実性を持つ一方、未学習のデータに対して予測が外れている状態）を示している。

## 4-3　原因分析

原因分析のため、4-1のモデル学習の際に利用したデータと予想するデータの関係性を図16に示すように整理した。

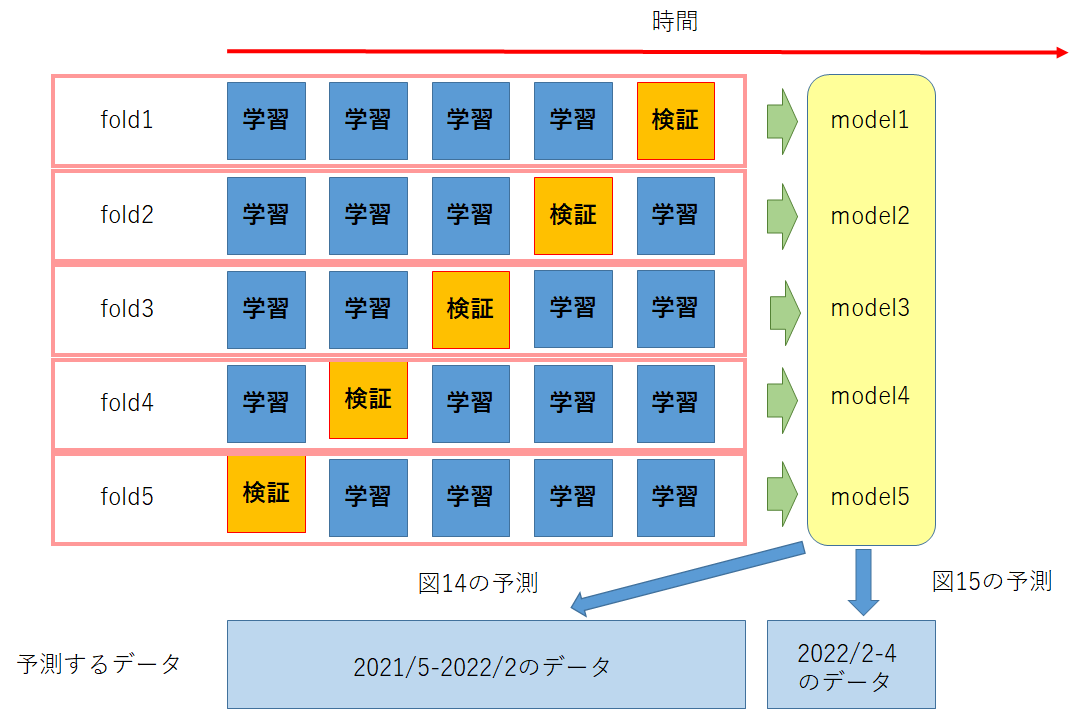


図16. 1st Trialモデルと予測データの関係

図16から、２つの問題点が浮かび上がった。図14の予測では、予測する対象を別のモデルで学習に使っていることである。例えば、fold3で検証に用いた左から3番目のデータグループは、他のfold（1,2,4,5）では学習に使用している。予測すべきデータを学習に利用しているため、すでに学習したデータの予測精度は良くなるが、未知のデータに対して予測精度が低くなる過学習が起こり易い。もう一つは本来知りえない未来のデータを使ってモデルを学習していることである。具体的にはfold1以外では未来の学習データを使って、過去の検証データの評価をおこなっている点である。過去と未来が分かっている状態では、現在のフォトマスク寸法の予測が良くなりすぎる現象が起きる。これはデータのリークとよばれる現象で、評価段階ではモデルの性能化過大に評価されるが、実際に運用する段階になると性能が上がらない原因となる。機械学習ではこのようなリークが起きないようにデータを準備することが必要である。このため、図17にデータのリークを避けた検証データの作成方法を考案した。



図17. リークを回避した学習データと検証データの分割方法（時系列データ）

図17では全データを9個のサブグループに分けて、検証データより過去の学習データだけを使って学習する方法が示されている。このような検証データを用いることで、より一般的で現実の運用に適したモデルを作成することができる。

このように検証データを準備して再度機械学習モデルを構築する。

**4.**実験結果

## 4-1　モデルの選定

　当初はLightGBMを選ぶという話、パラメータ調整したら似たような性能が出ることを確認

**参考文献**

1. **CS231n: Convolutional Neural Networks for Visual Recognition**

<http://cs231n.stanford.edu/>

1. **Géron, Aurélien. "Hands on Machine Learning with scikit-learn and Tensorflow." (2017).**

**Appendix**

1. フォトマスクに光を照射しウェア上に結像させる技術 [↑](#footnote-ref-1)