



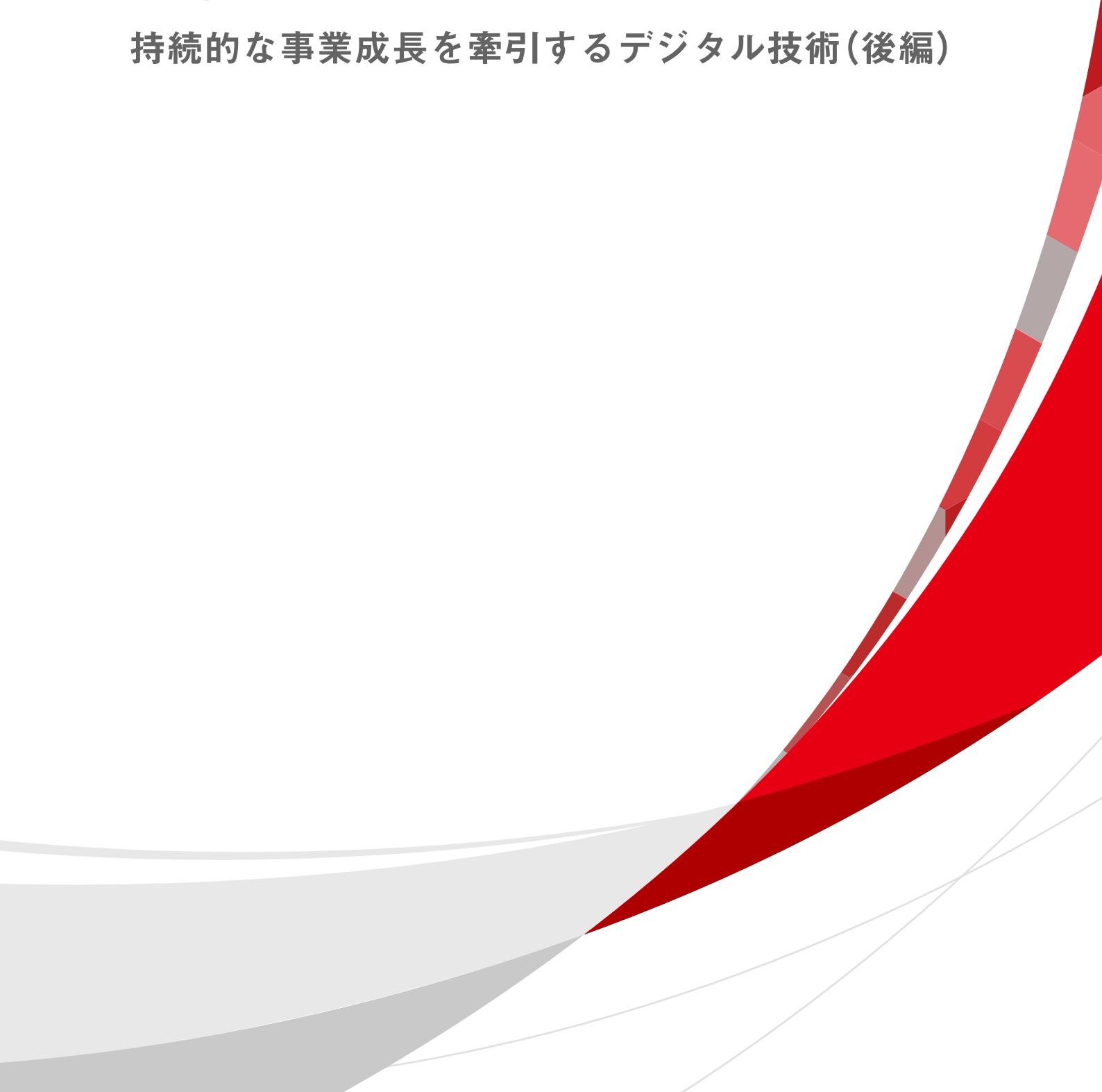
三菱電機技報

9

2025

Vol.99 No.9

持続的な事業成長を牽引するデジタル技術(後編)



No.9

特集 持続的な事業成長を牽引するデジタル技術(後編) Digital Technologies for Sustainable Business Growth (The second part)

特集論文

機械学習による材料・プロセス設計技術の深化と エポキシ系接着材料への適用	1-01
安藤圭理・加茂芳幸・坂東龍司・村上浩章・和田敏裕	
製造業ドメイン特化の言語モデル	2-01
内出隼人・齊藤辰彦・追立真吾・平森将裕・田口進也	
LLMを用いたラダープログラム の処理内容説明の生成技術	3-01
菅原直樹	
AI活用によるLSI開発の高度化	4-01
森 敦弘・高品信秀・南口比呂・深井哲人	
製造プロセスへのAI適用	5-01
氏家達浩・横堀祐芽・高阪 翔・榎原正朗	
生成AIを活用した社内技術文書検索システム の構築と展開	6-01
石橋祐太郎・宮下修治・深井哲人・中野智晴	
DevOps×生成AI×セキュリティーを搭載した ソフトウェア開発基盤“MelSecOps”	7-01
尾崎紀之・小森裕之・梅原友幸・長峯 基	
量子暗号	8-01
鶴丸豊広	
次世代データセンターの大容量化／低電力化 を実現する光電融合技術	9-01
白尾瑞基・大畠伸夫	

Advancements in Material and Process Design through Machine Learning: Application to Epoxy-Based Adhesive Materials	
Keisuke Ando, Yoshiyuki Kamo, Ryuji Bando, Hiroaki Murakami, Toshihiro Wada	
Domain-Specific Language Models for Manufacturing Industry	
Hayato Uchide, Tatsuhiko Saito, Shingo Odate, Masahiro Hiramori, Shinya Taguchi	
LLM-Based Program Analysis for Ladder Logic	
Naoki Sugawara	
AI-driven Enhancement in LSI Development Technology	
Atsuhiro Mori, Nobuhide Takashina, Hiro Minamiguchi, Yoshito Fukai	
AI Application to Manufacturing Processes	
Tatsuhiro Ujiiie, Yuka Yokobori, Kakeru Kosaka, Masaaki Sakakibara	
Building and Deploying Internal Technical Document Search System Using Generative AI	
Yutaro Ishibashi, Shuji Miyashita, Yoshito Fukai, Tomoharu Nakano	
Development Infrastructure Equipped with DevOps, Generative AI and Security “MelSecOps”	
Motoyuki Ozaki, Hiroyuki Komori, Tomoyuki Umehara, Motoi Nagamine	
Quantum Cryptography	
Toyohiro Tsurumaru	
Opto-electronic Convergence Technology for Achieving High Capacity and Low Power Consumption in Next-generation Data Centers	
Mizuki Shirao, Nobuo Ohata	

執筆者の所属は執筆時のものです。

本号に記載されている会社名、製品名はそれぞれの会社の商標又は登録商標です。

三菱電機では、サステナビリティ経営を実現するビジネスエリアとして、「インフラ」「インダストリー・モビリティ」「ライフ」を設定しています。

三菱電機技報ではこのビジネスエリアを中心に特集を紹介しています。

今回の特集では全エリアを支える基盤となる“持続的な事業成長を牽引するデジタル技術(後編)”をご紹介します。

特集論文

機械学習による材料・プロセス設計技術の深化とエポキシ系接着材料への適用

Advancements in Material and Process Design through Machine Learning: Application to Epoxy-Based Adhesive Materials

安藤圭理*

Keisuke Ando

加茂芳幸†

Yoshiyuki Kamo

坂東龍司*

Kyuji Bando

村上浩章*

Hiroaki Murakami

和田敏裕*

Toshihiro Wada

*先端技術総合研究所

†同研究所(博士(工学))

要 旨

電気機器の高性能化や高信頼化、環境負荷低減を支える材料技術の開発では、原料の化学組成や製造プロセスなどの設計パラメーターを調整する必要がある。しかし、設計パラメーターの変更によって材料の微細な構造や状態が劇的に変化するため、望ましい構造や状態を持つ所望の材料の開発には試行錯誤に膨大な時間を要する。

そこで、既知の実験データを用いた機械学習によって有望な設計パラメーター範囲を推定した上で、予測の不確実性が高く、かつ実験データの少ない条件を次の実験点に選定することで、有望な設計パラメーター範囲を効率良く明確化する手法を開発した。さらに、この手法をエポキシ系接着材料の開発に適用し、所望の材料が得られる原料組成比の明確化に必要な実験回数を従来比約80%削減した。この技術は新規材料の開発だけでなく、既存材料や製造プロセスなどの改善といった様々な用途への適用が期待できる。

1. まえがき

材料技術は、幅広い産業課題・社会課題を解決に導く可能性のある分野横断的な基盤技術である⁽¹⁾。近年はマテリアルズインフォマティクス(MI)を取り入れた材料技術の研究開発が国内外の産官学で盛んに取り組まれている⁽¹⁾⁽²⁾。三菱電機では、家電から宇宙までのあらゆる環境で価値を提供する電気機器製品を製造しており、高性能化や高信頼化、環境負荷の低減など、電気機器の更なる高付加価値化を実現する材料開発が求められている。材料開発では、候補になる材料を創出した上で、さらに機器で実用できるように原料の化学組成や製造プロセスなどの設計パラメーターを調整する必要があるが、わずかな条件の違いが材料の微細な構造や状態に劇的な変化をもたらす。そのため、当初から所望の材料が得られる有望な設計パラメーターの範囲を把握できれば、効率的で革新的な開発が期待できる。

今回、新規エポキシ系接着材料に使用するエポキシモノリスシートを具体的な例として、MIによって既知の実験結果から能動的に有望な設計パラメーターの範囲を効率良く明確化する手法を検討した。エポキシモノリスシートは、3種の原料の組成比によって微細構造が変化し、またエポキシモノリスシートが得られない原料組成比も存在する。さらに、エポキシモノリスシートの形成は、原料の反応性にも強く影響されるため、原料の化学構造の違いにも依存する。そのため、エポキシモノリスシートが得られる原料組成比などの設計パラメーター範囲を特定するには、実験的な試行錯誤を要して膨大な時間が必要であった。

本稿では、既知の実験データを用いた機械学習によって有望な設計パラメーター範囲を推定した上で、予測の不確実性が高く、かつ実験データの少ない条件を次の実験点に選定することで、有望な設計パラメーターの範囲を効率良く明確化するMI手法を開発し、エポキシモノリスシートに適用した事例について述べる。

2. 有望な設計パラメーターの範囲の推定方法

今回の目的のように、複数の原料の組成比などの条件が可変であるという前提の下、所望の材料が得られるパラメーター範囲を明確化する手法として、能動学習によって材料の温度や成分に応じた安定状態を示す相図を推定する取組みがある。Daiらは、ガウス過程を用いた事後予測に基づいたベイズ最適化を用いることで、二つの相を持つ相図を推定する過程で必要なサンプリング点を大幅に削減できることを明らかにした⁽³⁾。Terayamaらは、相境界付近の不確かな領域を重点的に探索する不確実性サンプリングによって、複数の相を持つ相図でも効率的なサンプリングが可能であることを明らかにした⁽⁴⁾。これら先行研究が対象とする相図推定では、熱力学的平衡状態を取り扱うため、熱力学計算を活用できる場合も多くあることから、ある程度多いサンプリング回数も許容できる。一方、本稿が対象とするエポキシモノリスのよ

うに、製造プロセスに応じて微細構造が変化することを特徴とする材料では、実験的な検証が必要であるため、サンプリング回数の制約が厳しい。そこで、効率良く有望な設計パラメーター範囲を推定するために、不確実性サンプリングで、これまで検証したことのある教師データとの距離が離れている条件を優先的に提案する手順を導入した。

2.1 推定フレームワーク

図1に、有望な設計パラメーター範囲の推定フレームワークを示す。ここでは、設計パラメーターとして3種の原料A, B, Cの比率を選択した場合を例示した。図1の三角図は原料組成比を表しており、原料Aの頂点に近ければ、原料Aが多いことを意味している。

まず、初期の教師データとして、所望の材料が得られた結果と望ましくない結果の2種類を用意する。ここで、必要な原料が欠損している組成比では、望ましくない結果が得られることは明らかであることから、三角図の頂点では望ましくない結果が得られるなどを教師データに追加した。

次に、この教師データを用いて機械学習を行うことで、設計パラメーターが有望であるかどうかを分類する予測モデルを構築し、有望な設計パラメーター範囲を推定した。この推定結果に基づいて不確実性スコアを評価するとともに、罰則項として教師データとの近さを算出し、それによってデータ希薄性スコアを評価した。不確実性スコアは、設計パラメーターが有望であるかどうかを予測モデルが分類する不確かさを評価する数値であり、有望と推定された範囲と、望ましくないと推定された範囲の境界が最も不確実であると算出されることから、境界の全周が同等に不確実であると評価される。そこで、不確実性スコアにデータ希薄性スコアを加えた評価関数を設計することで、不確実であると同時にこれまで実験を行っていない候補条件を一義に提案可能にした。この手法で提案された候補条件で実験的に検証を行い、所望の材料が得られるか確認した。そしてこの結果を教師データとして追加して推定・評価・検証のサイクルを繰り返すことで、有望な設計パラメーター範囲を明らかにした。

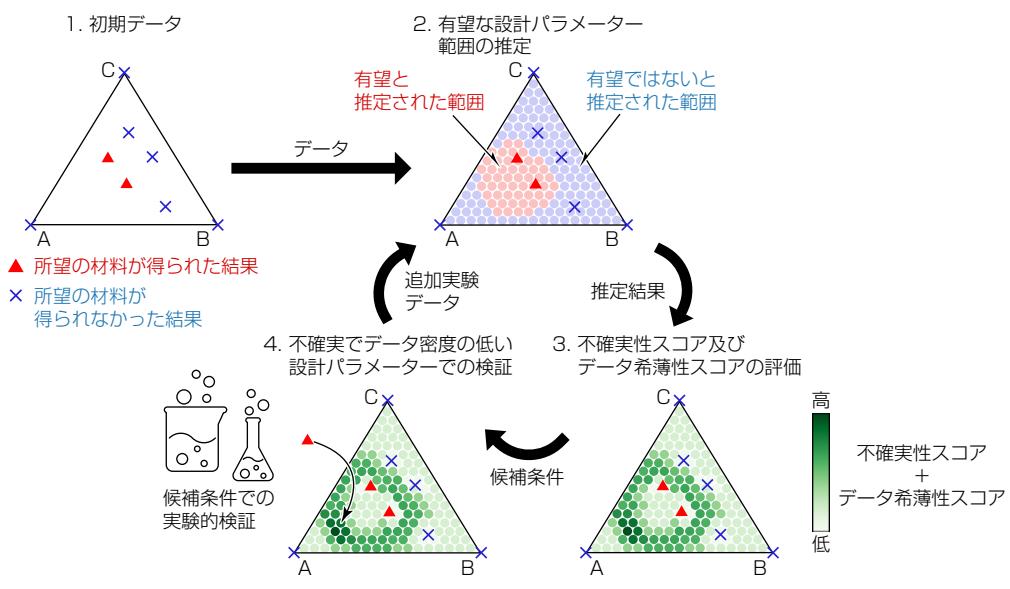
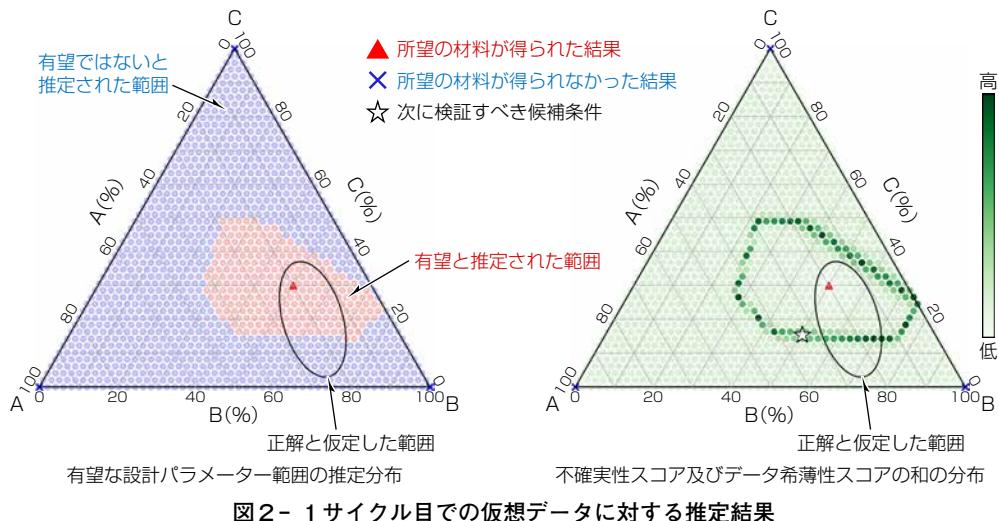


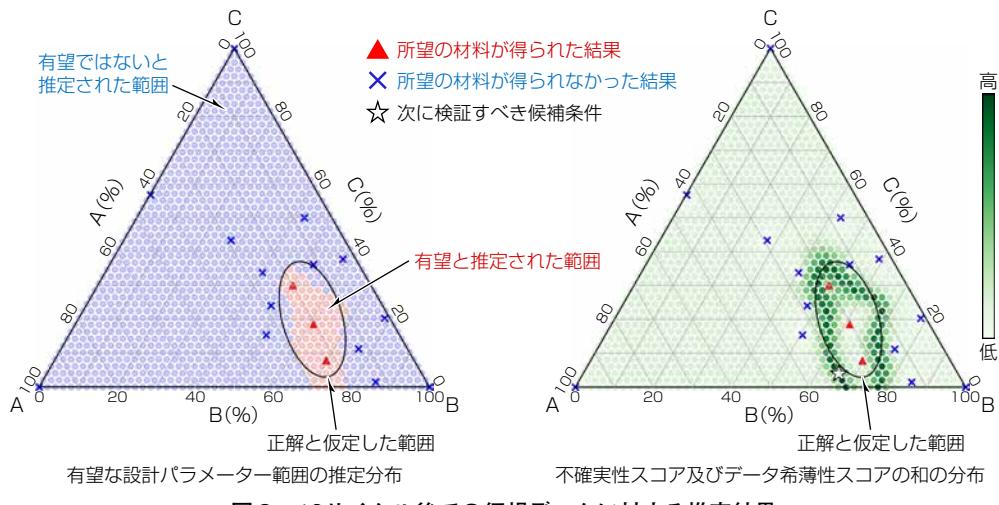
図1-有望な設計パラメーター範囲の推定フレームワーク

2.2 推定フレームワークの事前検証

この手法の動作を確認するために、仮想的な正解を設定し、有望な設計パラメーター範囲を推定できるかを検証した。図2に、1サイクル目の推定結果を示す。図中の黒線で示す楕円(だえん)の内部が正解と仮定した範囲であり、図2左の赤い領域が有望と推定された範囲を意味する。1サイクル目では推定された範囲と正解は一致していないことが分かる。図2右では、推定結果から算出された不確実性スコア及びデータ希薄性スコアの和の分布を示している。有望と推定された範囲と、望ましくないと推定された範囲の境界で、不確実性スコアとデータ希薄性スコアの和が高く評価されており、星印が次に検証すべき候補条件として提案される。候補条件が正解と仮定した範囲内であれば所望の材料が得られたという仮想データを追加し、範囲外であれば所望の材料が得られなかったという仮想データを追加した。



推定フレームワークを13サイクル繰り返した後の推定結果を図3に示す。13回の仮想データの追加によって、有望と推定された範囲は正解に非常に近い結果を示しており、少ない実験でも有望な設計パラメーター範囲を導ける可能性を確認した。



3. 新規エポキシ系接着材料の開発への適用

この章では、新規エポキシ系接着材料に使用するエポキシモノリスシートにこの手法を適用した事例について述べる。

3.1 新規エポキシ系接着材料に使用するエポキシモノリスシート

電気モーターや半導体デバイス、航空宇宙機器に広く使用されるエポキシ系接着材料は、高強度で耐熱性に優れた材料である一方、塑性変形しにくく脆(もろ)いという欠点もある。電気機器では、動作時と停止時の温度サイクルによって周辺部材が膨張と収縮を繰り返すため、エポキシ系接着材料の柔軟性の向上が求められる。当社では、連続気孔を持ったエポキシモノリスシートに接着成分を含侵させたシート接着材料を開発し、柔軟なエポキシモノリスシートを接着接合部の応力緩和層として機能させることで、エポキシ系接着材料の温度サイクル耐性向上に寄与することを明らかにしており⁽⁵⁾、更なる高性能化に向けて研究開発を進めている。

エポキシモノリスシートは、網目状のエポキシ骨格と空隙がそれぞれ三次元的につながった連続多孔体構造を特徴とする薄膜である(図4)⁽⁶⁾。エポキシモノリスシートを実現するためには、主剤と硬化剤、細孔形成剤の少なくとも3種の原料を混合する必要があり、原料組成比の違いによって微細組織は変化する。原料組成比が適切な範囲内ではないと、エポキシ骨格が不連続になってシートの自立性を確保できず、又は、空隙が不連続になって接着成分を含侵できず、接着材料として成立しない。このことから、エポキシモノリスシートが得られる原料組成比の明確化が求められていた。

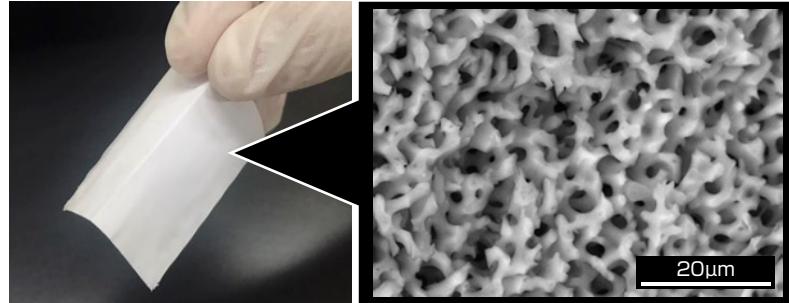


図4-エポキシモノリスシート

3.2 エポキシモノリスシートを得るために有望な設計パラメーター範囲の推定方法

本稿では、硬化剤に4, 4'-メチレンビス(シクロヘキシルアミン)(BACM)、細孔形成剤にポリエチレングリコール(PEG)を用いて、主剤として2, 2'-ビス(4'-グリシジルオキシフェニル)プロパン(BADGE)、又は1, 3-ビス(N, N-ジグリシジルアミノメチル)シクロヘキサン(TETRAD-C)を用いたエポキシモノリスを対象とした。この二つのエポキシモノリスの系に対して、それぞれ原料3種の組成比を設計パラメーターに設定し、エポキシモノリスシートが得られる有望な設計パラメーター範囲を推定した。まず、事前の実験でエポキシモノリスが形成されたPEG濃度70wt%の結果と、モノリスが形成しない各原料の単体の結果(三角図の頂点)を初期の教師データとして使用した。2. 2節での事前検証に基づいて、教師データに追加実験の結果を加えつつ、推定フレームワークを13サイクル繰り返した。

4. エポキシモノリスシートが得られる有望な設計パラメーター範囲の推定結果

図5に、13サイクル後に推定された、エポキシモノリスが得られる設計パラメーター範囲を示す。実験を行った条件の分布に注目すると、2. 2節での事前検証の結果と同様、モノリスの作製に有望と推定された範囲の全周を囲うように、モノリスの作製に有望ではない実験点が分布していることから、有望な設計パラメーター範囲を高精度に推定できていると考えられる。特に、主剤にBADGEを用いたモノリスとTETRAD-Cを用いたモノリスとで有望な設計パラメーター

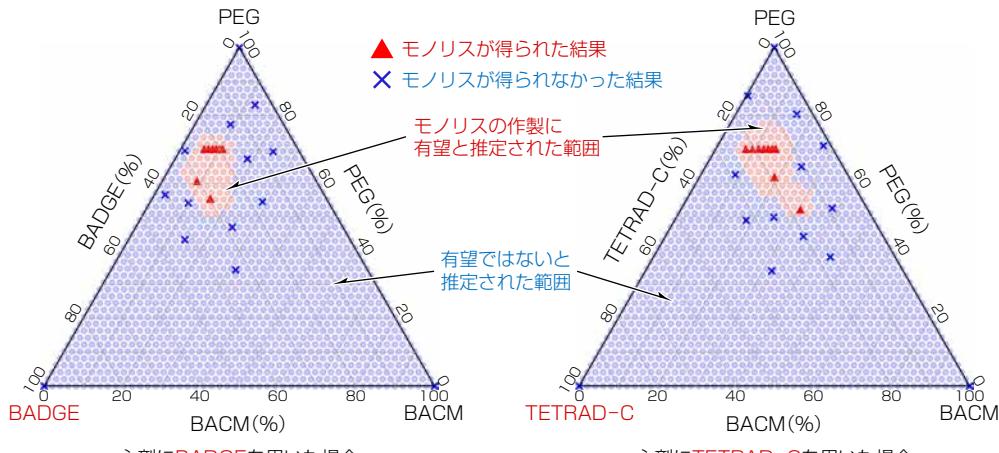


図5-エポキシモノリスシートが得られる設計パラメーター範囲

範囲が異なっていることから、原料の種類の違いを反映した設計パラメーター範囲を推定できている。このように有望な設計パラメーター範囲を明確化できたことから、少ない実験でエポキシモノリス合成の全体的な傾向を把握した上で材料を設計できるようになった。また、13サイクル後の推定結果でエポキシモノリスの作製に有望ではないと推定されたPEG濃度40wt%以下の実験回数を減らして、有望である条件での実験に注力できた。従来、10wt%刻みで全条件の実験を行う場合には66回の追加実験が必要であったのに対して、この手法によって13回の追加実験で有望な設計パラメーター範囲を明らかにできた。実験回数を約80%削減しており、この手法が研究開発の効率化に有効であることを示した(図6)。

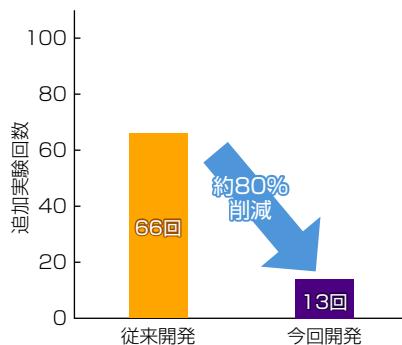


図6-有望な設計パラメーター範囲の明確化に必要な実験回数の変化

5. むすび

革新的な材料の効率的な研究開発のために、機械学習を用いて設計パラメーターを効率的に選定する手法を開発した。多数の設計パラメーターを含む系でも、設計パラメーター範囲を少ない実験回数で高精度に推定する手法を見いだして、この手法を新規エポキシ系接着材料の開発に適用することで、従来開発に比較して、約80%の実験回数を削減した。この手法は設計パラメーターの種類が三次元以上の高次元に対しても拡張可能であるため、今後は更に複雑で高度な材料開発にも適用を進める。

参 考 文 献

- (1) 内閣府 マテリアル戦略有識者会議：マテリアル革新力の一層の強化に向けた論点と検討の方向性 (2024)
<https://www8.cao.go.jp/cstp/material/11kai/siryol-1.pdf>
- (2) 国立研究開発法人 科学技術振興機構 研究開発戦略センター：研究開発の俯瞰報告書 ナノテクノロジー・材料分野 (2024年)
<https://www.jst.go.jp/crds/report/CRDS-FY2024-FR-04.html>
- (3) Dai, C., et al. : Efficient Phase Diagram Sampling by Active Learning, *The Journal of Physical Chemistry B*, **124**, 1275~1284 (2020)
- (4) Terayama, K., et al. : Efficient construction method for phase diagrams using uncertainty sampling, *Physical Review Materials* 3, 033802-1~033802-8 (2019)
- (5) Kamo, Y., et al. : Stress relaxation and improved fracture toughness of metal bonding using flexible monolith sheets and an epoxy adhesive, *Polymer Journal*, **57**, 203~214 (2025)
- (6) 松本章一：エポキシモノリスを用いる異種材料接合, *ネットワークポリマー*, **38**, No.2, 93~102 (2017)

製造業ドメイン特化の言語モデル

Domain - Specific Language Models for Manufacturing Industry

内出隼人*

Hayato Uchide

平森将裕†

Masahiro Hiramori

斎藤辰彦*

Tatsuhiiko Saito

田口進也*

Shinya Taguchi

追立真吾†

Shingo Oidate

*情報技術総合研究所

†設計技術開発センター

要 旨

大規模言語モデル(Large Language Model : LLM)の進展によって、自然言語処理技術(Natural Language Processing : NLP)の実用化が急速に進んでいる。一方、モデルの大規模化に伴う計算コストやエネルギー消費の増大、データプライバシーの懸念、リアルタイム応答性の制約など、製造現場で運用する上で様々な問題が顕在化している。特に、生産ライン停止時の原因特定や対処手順の即時提示といった、生産性に直結する緊急性の高い場面では、限られた計算環境でのモデルの安定動作と高い応答性が求められる。

そこで三菱電機は、FA分野をはじめとする製造業ドメインに特化した小規模言語モデル(Small Language Model : SLM)を開発した。今回、継続事前学習、指示チューニングに加えて、限定されたデータ環境下でも効果的に学習できるアライメントを新たに検討した。その結果、パラメーター数18億個という、エッジデバイス上で動作可能なコンパクトなモデルでありながら、FA分野の知識の正誤を問うタスクで、正解率77.24%を実現できた。これによって、製造現場などの制約がある環境での生成AI活用範囲の拡大が期待できる。

1. まえがき

LLMの進展によって、NLPの実用化が急速に進んでいる。特に、OpenAIのChatGPT^(注1)をはじめとする、クラウド上で動作する汎用LLMは、幅広い言語処理タスクで高い性能を発揮し、製造業を含む多くの産業領域での活用が期待されている。一方、LLMの活用で、モデルの大規模化に伴う計算コストやエネルギー消費の増大が問題になっている。

また、運用の場面では、クラウド上の汎用LLMを単純に利用するだけでは解決が難しい場合がある。例えば、製造現場では、設備マニュアルやトラブル対応記録などの機密情報を外部に送信することが困難な場合や、現場の作業者や設備と即時性の高いやり取りが求められる場合が多く存在する。具体的には、生産ライン停止時の原因特定や対処手順の即時提示、保守作業中の確認支援、異常発生時の対応判断が挙げられる。これらの場合は、設備マニュアルやトラブル対応記録などの機密情報に基づいた、即時性の高い情報取得・応答が、生産性や安全性に直結する。こうした背景から、クラウド上の汎用LLMに依存せずに運用できる仕組みが求められている。特に、セキュアで応答性の高い処理を現場端末で完結させるためには、エッジデバイス上で動作する高性能なSLMの技術開発が重要である。

SLMの開発・運用で、単にモデルサイズを抑えるだけでは十分な性能を得ることは難しい。特に製造業のように専門性の高いドメインでは、汎用モデルには含まれない技術用語や表現が多い。専門性の高いドメインで、ユーザーからの問合せや現場の状況に応じた適切な応答を生成するためには、適切なドメイン適応が不可欠である。そのため、特定ドメインのデータを活用した継続事前学習⁽¹⁾や指示チューニング⁽²⁾、さらにユーザーにとって自然かつ安全な応答を実現するためのアライメント⁽³⁾といったファインチューニング手法を適切に組み合わせる必要がある。

また、製造現場などの制約がある環境で言語モデルを実用的に動作させるためには、推論精度を確保しつつ、応答速度やリソース使用量とのバランスを考慮した最適化も重要になる。特に即時性が求められるユースケースでは、現場での応答遅延を最小限に抑える設計や、限られた計算環境で安定動作を実現する工夫が、モデルの性能と実用性を両立する上で重要である。

本稿では、製造業の中でも特にFA分野に特化し、当社FA製品に関する各種ドキュメントを用いてSLM(図1)を構築した取組みを述べる。2章では、代表的なファインチューニング手法と、それらを用いたドメイン特化の実践例を述べる。3章では、OSSを活用した、エッジデバイス上でSLMを動作させる実践例を述べる。最後に4章では、今回の成果と今後の展望についてまとめる。

(注1) ChatGPTは、OpenAI OpCo, LLCの登録商標である。

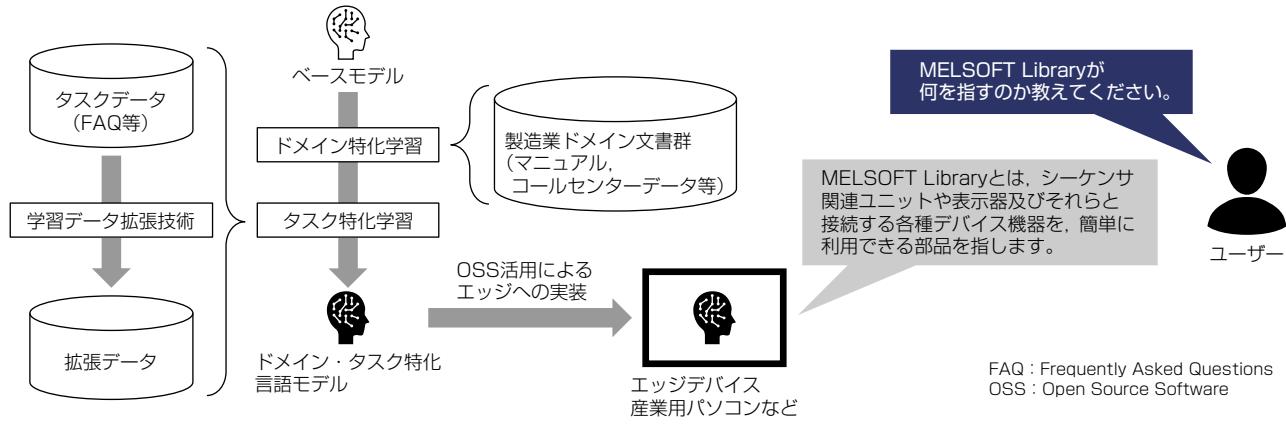


図1-エッジデバイスで動作する製造業ドメインに特化したSLM

2. 製造業ドメイン特化のSLMの開発

LLMは、大量の一般的なテキストデータを用いて事前学習されることによって、幅広い言語処理タスクへの高い汎用性を獲得する。一方、実際の応用では、特定の言語、ドメイン、利用目的に適合するようにモデルを追加で調整する必要がある。こうした目的のための事後学習には、幾つかの代表的な手法があり、それぞれの目的や適用対象に応じて使い分けられる。

この章では、言語モデルの性能を最大限に引き出すための代表的なファインチューニング手法を述べて、今回の開発で取り組んだ製造業ドメイン特化手法を述べる。

2.1 繼続事前学習

継続事前学習は、既存の事前学習済みモデルに対して、新たなコーパスを用いて再度事前学習する手法である。これによって、特定言語(例：日本語)やドメイン(例：製造業、医療、法律)での語彙や表現、文脈の知識をモデルに追加的に学習させる。

継続事前学習は自己教師あり学習に基づくため、明示的なタスク定義やアノテーションが不要であるという利点があり、社内文書や設備マニュアル、トラブル対応記録のような非構造データを活用しやすい。

特に日本語や専門性の高いドメインでは、汎用LLMが十分にカバーしていない言語現象や語彙、表記揺れが多数存在するため、継続事前学習は有効な手段になる。一方、既存モデルの性能を劣化させないよう、事前学習データとのバランス設計やトークナイザーの調整が重要になる。

2.2 指示チューニング

指示チューニングは、モデルに“タスクを理解し、指示に従って応答する能力”を学習させるための手法である。具体的には、“命令(instruction)”とそれに対応する“応答(response)”のペアを用いて教師あり学習を行う。例えば、“次の文章を敬語に書き換えてください”といった指示に対して、意味は変えずに自然な敬語の文章を生成するように学習させる。

この手法は、チャット形式やプロンプト形式のインターフェースで、ユーザーが直感的にモデルを活用するために不可欠であり、LLMの対話的な利用を可能にする上で大きな役割を果たす。また、多様なタスクへの汎用性を高めるために、様々な形式の命令データを収集・構築する工夫が求められる。

指示チューニングは、英語で豊富なデータが存在する一方、日本語では高品質なデータセットが限られており、翻訳データの活用や日本語特有のタスク設計が重要になる。

2.3 アライメント

アライメントは、モデルの応答をユーザーの価値観や社会的規範、安全性の観点に整合させるための手法群を指して、単に“正しい”応答を出力するだけでなく、“ユーザーにとって望ましく適切な”応答を生成することを目指す。また、応答

の品質だけでなく、安全性や倫理的な配慮が重要視される。アライメントの実現には、主に人間の選好に基づいた学習と安全性チューニングという二つのアプローチがある。

人間の選好に基づいた学習では、人間が評価した“より好ましい応答”と“望ましくない応答”的ペアを用いることで、モデルがユーザーにとって適切な応答を選択する傾向を学習する。この方法は、モデルの出力を評価し、より高い品質と整合性を持つ応答を強化するために利用される。例えば、DPO(Direct Preference Optimization)⁽⁴⁾やPPO(Proximal Policy Optimization)⁽⁵⁾といった手法が広く用いられており、これらは報酬信号を通じてモデルの出力を調整することで、応答品質を向上させることを可能にする。

また、安全性チューニングは、モデルが攻撃的、不適切、又は危険な内容を含む出力を抑制することを目的とする。特に、日本語を対象とする場合、英語とは異なる文化的背景や社会的規範を考慮した安全性の確保が必要になる。例えば、日本語の言語的特性や社会的慣習を考慮したフィルタリング基準を設けることで、意図しない誤解や不適切な応答を回避することが求められる。このような取組みは、モデルを業務システムや公共のサービスに組み込む際の信頼性向上に直結するものであり、特に製造業のように安全性が重視されるドメインでは欠かせない要素である。

このように、アライメントのプロセスは、モデルの応答をユーザーの期待や要求に適合させるための重要な手段である。適切なアライメントを実現することによって、モデルの信頼性と実用性を高めることができて、より安全で効果的なシステムの構築が可能になる。

2.4 製造業ドメイン特化手法

ここまで述べた手法を用いた、製造業ドメイン特化手法を述べる。製造業ドメイン特化手法のフローを図2に示す。今回の取組みでは、当社FA製品に関する各種ドキュメントを使用し、継続事前学習、指示チューニング、アライメントの三つの手法を適用した。これによって、製造業特有の専門知識や用語を効率的に学習させて、モデル性能の向上を図った。

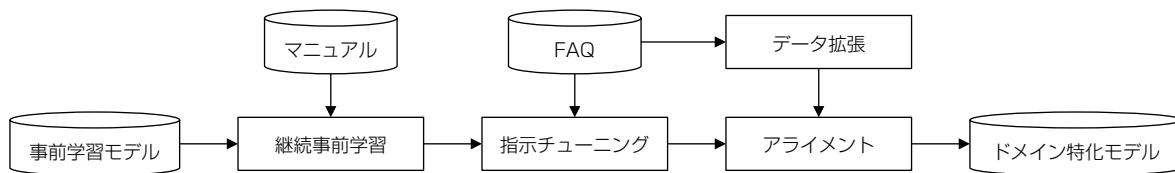


図2-製造業ドメイン特化手法のフロー

ベースとなる事前学習モデルには、18億個のパラメーター数のモデル⁽⁶⁾(モデルアーキテクチャー: Llama2⁽⁷⁾)を採用した。モデルサイズの選定では、製造業ドメインに特化した学習を効率的に行うために、計算コストと性能のバランスを考慮した。特に製造業では、リソース制約を踏まえつつ高精度なモデルを構築することが求められるため、この規模のモデルを選定した。また、当社ではモデルの学習データを適切に扱って、将来的にモデルの出力結果に対して説明できるよう、モデルの透明性を高めることを重要視している。この観点から、学習データも併せて公開されているこのベースモデルを採用するに至った。それ以降の処理では、このベースモデルを用いて各種ドメイン特化手法を適用した。

継続事前学習では、FA製品に関するマニュアルデータを使用した。このマニュアルデータは、製品の仕様、操作手順、トラブルシューティングなど幅広い内容を含む。事前学習済みモデルに対して、この専門的なコーパスを再学習させることで、FA製品に特有の語彙や表現、構造化された情報を効率的に取り込むことを目指した。

指示チューニングでは、FA製品に関するFAQデータを利用した。このデータは製品の使用方法やトラブル対応に関する質問とその回答を含む形式で構成されている。指示チューニングによって、モデルはユーザーからの質問に対して適切な応答を生成する能力を向上させることができる。特に、質問形式での入力に対して自然な応答を提供するためには、この段階でのチューニングが不可欠である。さらに、製造業に特有の問合せに対応するため、FAQデータの内容も精査し、適切な形式に整備した。

アライメントでは、指示チューニングと同様に、FA製品に関するFAQデータを活用した。アライメントの目的は、モデルがユーザーにとって望ましい応答を生成する能力を高めることである。しかしながら、FAQデータは質問とそれに対応する“適切な回答”が含まれている一方、“望ましくない応答”が存在しない点に対応する必要があった。このため、既存のFAQデータ内の回答テキスト群から、テキストの類似度が高い、似て非なる回答テキストを抽出し、これを同一質問に対する“望ましくない応答”とみなすデータ拡張手法を新たに検討した。この手法によって、人手による追加アノテー

ションや“望ましくない応答”的設計を行わずに、既存のドメインデータから効率的にアライメント用のデータペアを用意できるようになった。

こうした継続事前学習、指示チューニング、アライメントの手法を組み合わせることで、製造業ドメインに特化したモデルの構築を実行可能にした。モデルの性能評価は、評価モデルにAnthropic Claude-3.7 Sonnetを用いたLLM-as-a-judge⁽⁸⁾による参考回答付きの自動評価を採用した。評価データには、当社FA製品に関する知識の正誤を問う質問群を使用し、比較対象には代表的なクラウド上の汎用LLMであるOpenAI GPT-4oを選定した。評価データに対する、今回開発したモデルの応答と、比較対象であるGPT-4oの応答について、参考回答と比較し正誤を判定させた。この比較によって、開発したモデルの有効性を、一定の客観性をもって評価できる。

評価結果として、今回開発したモデルは、正解率77.24%になることを確認した。比較対象であるGPT-4oは正解率52.03%であった。この評価は当社FA製品に関する知識の正誤を答える2択問題であるため、ランダムに答えた場合の正解率(チャンスレート)は50%が期待される。GPT-4oの正解率はチャンスレートに近い結果であった。一方、今回開発したモデルはチャンスレートを27.24%上回る結果であり、この開発で検討したドメイン特化手法の有効性を示すことができた。

3. OSS活用によるエッジAIへの実装

ドメイン特化SLMは、クラウド上でLLMを動作させる従来のアプローチとは異なり、軽量で、かつ高い応答性が求められる現場のニーズに応えるため、エッジデバイスでの実行(エッジAI)が期待されている。エッジAIとしてSLMを動作させることで、低遅延かつプライバシーを考慮した処理が可能になり、スマートファクトリーやエッジロボティクス、エネルギー制御など多様な分野で、その価値は大きい。

エッジデバイス上でドメイン特化SLMを効率的に動作させるには、限られたハードウェアリソースの中で性能・精度を両立させる最適化が不可欠である。最適化の手段として、柔軟かつ高機能なOSSの活用が重要になる。当社は、AI開発の基盤となるOSSのコミュニティーで、世界トップクラスの技術者とともに開発に参画し、技術力の向上と社会への貢献を同時に果たしている。具体的には、AIコンパイラOSSであるApache TVM^{(注2)(9)}で、当社のエンジニアがソースコードの編集権限を持つ“コミッター”として中核的な立場で活動しており、またAIフレームワークとして事実上の業界標準となっているPyTorch^(注3)でも当社の貢献が評価されて、PyTorch Contributor Awards 2024⁽¹⁰⁾の最終候補者として選出された。

この章では、これらOSSに関する知見を生かして、当社が取り組んだSLMをエッジデバイス上で動作させる実証について述べる。対象デバイスは、GPUを搭載したJetson^(注4) Orin Nano 8 GBと、NPU(Neural Processing Unit)を搭載したRadxa ROCK 5B 16GBの2種であり、どちらも産業用途での応用を視野に入れた選定である。今回は、この両デバイス上で37億個のパラメーター数のSLM(モデルアーキテクチャー：Llama2)をOSS活用によって省メモリーかつ高速に実行する方式を検討した。検討に先立って、PyTorchのEagerモード実行といった一般的な実行方法では対象デバイスのメモリー不足(必要メモリー量は14.9GB)のため前記SLMが実行不可能であることを確認し、この取組みの重要性を明確にした。

エッジ上でSLMを実行するに当たって、まず適切なLLM推論用OSSを選定する必要がある。OSS選定に当たっては、2章で述べたSLM(モデルアーキテクチャー：Llama2)に対応していること、対象デバイスに対応していること、の2点を満たすOSSを候補として、OSS活用の有用性を各々評価した。

Jetson Orin Nano 8GBでは、MLC-LLM(Apache TVMベース)、ExecuTorch、ollama、vLLM、IREE、llama.cpp⁽¹¹⁾が選定候補になり、最終的にllama.cppを採用した。llama.cppは、推論の実装方針として“handcrafting方式”すなわちハードウェアごとに手動でコードを最適化するスタイルを取っており、開発初期段階で柔軟性が高く、導入が比較的容易である点を重視して採用した。これによって、Jetson Orin Nano 8 GB上でSLM実行が可能になり、実用的な推論速度を確認できた。実行時の生成速度とメモリー使用量については表1に示す。

Radxa ROCK 5B 16GBでは、MLC-LLM(Apache TVMベース)、rkllmが選定候補になったが、MLC-LLMは対象デバイスへの対応が十分ではなかったため、rkllmを採用した。実行時の生成速度とメモリー使用量については表2に示す。

このように、二つの対象デバイスに対して最適なOSSを活用することで、一般的なAI実行方法ではメモリー不足によって実行不可能であった37億個のパラメーター数のSLMが実行可能であることを確認した。

表1-エッジデバイスでのドメイン特化SLM実行結果(Jetson Orin Nano 8GB)

対象デバイス	Jetson Orin Nano 8GB(GPU搭載)
実行対象SLM	3.7BクラスSLM(アーキテクチャー:Llama2)
OSS選定候補	MLC-LLM(Apache TVM), ExecuTorch, ollama, vLLM, IREE, llama.cpp
採用OSS	llama.cpp
生成速度	22トークン/秒
メモリー使用量	2.4GB

表2-エッジデバイスでのドメイン特化SLM実行結果(Radxa ROCK 5B 16GB)

対象デバイス	Radxa ROCK 5B 16GB(NPU搭載)
実行対象SLM	3.7BクラスSLM(アーキテクチャー:Llama2)
OSS選定候補	MLC-LLM(Apache TVMベース), rkllm
採用OSS	rkllm
生成速度	6.8トークン/秒
メモリー使用量	4.5GB

なお、LLM推論フレームワークの設計方針には大きく分けて二つの方式が存在する。一つは、今回採用したllama.cppのような“handcrafting方式”で、ソースコードを直接調整することで高い柔軟性を実現する方式である。もう一つは、Apache TVMに代表される“compiler方式”で、対象デバイスに合わせた最適化コードを自動生成する手法である。

handcrafting方式は、LoRA(Low-Rank Adaptation)のような効率的なファインチューニング手法への対応で柔軟であり、導入も比較的容易である。一方、対応可能なプラットフォームやデバイスは限定的であり、対象デバイスごとの実装・最適化が必要になるため、拡張性や保守性の面で課題が残る。

これに対してcompiler方式は、現時点でLoRAに対応に制約があるものの、MLC-LLMのようにモバイル環境やブラウザ環境を含む幅広いプラットフォームに対応しており、将来的な拡張性やスケーラビリティの観点で優れている。また、コード生成によって環境ごとの最適化を自動化できるため、長期的には保守性や開発生産性の向上が期待できる。

当社は、今後更に多様化するデバイス環境やユースケースへの対応力を強化するため、compiler方式への移行を進める方針である。その上で、LoRAをはじめとする学習最適化機能の実装については、OSSコミュニティと連携しながら対応機能の拡充を目指す。これまでに培ってきたOSS活用技術を生かして、エッジAI領域でのドメイン特化SLMの実用化と普及を加速し、更に持続可能で拡張性の高い社会の実現に取り組んでいく。

(注2) Apache TVMは、Apache Software Foundationの登録商標である。

(注3) PyTorchは、The Linux Foundationの登録商標である。

(注4) Jetsonは、NVIDIA Corp.の登録商標である。

4. む す び

今回の結果は、製造業のように専門的かつ限定的なドメインに特化したSLMの開発で、継続事前学習、指示チューニング、アライメントの組合せが有効であることを示している。また、この手法はほかの製造業関連ドメインや異なる業種に対しても応用可能であり、適用範囲を広げることで更なる性能向上が期待される。

今後の課題として、モデル性能を更に向上させるための追加データの収集や、アライメント手法の改善が挙げられる。特に、望ましくない応答の自動生成手法の精緻化や、応答の多様性と一貫性を両立させるための工夫が求められる。また、モデル評価方法についても、より多角的な指標を用いて実用性を定量的に測定する手法を検討する必要がある。それに加えて、多岐にわたるユースケースに対応するための、エッジデバイス上での最適化を検討する必要がある。ユースケースによっては、現状の想定よりも高い応答性や少ないリソース使用量が求められる可能性がある。そのため、限られた計算環境でのモデルの安定動作と高い応答性を実現するための、実装面の工夫が必要である。これらの課題に継続的に取り組むことで、製造業ドメインでのSLMの更なる高度化と実用化を実現したいと考えている。

参考文献

- (1) Gupta, K., et al. : Continual Pre-Training of Large Language Models: How to (re)warm your model?, The Thirty-Seventh Annual Conference on Neural Information Processing Systems (2023)
- (2) Wei, J., et al. : Finetuned Language Models are Zero-Shot Learners, The Tenth International Conference on Learning Representations (2022)
- (3) Bai, Y., et al. : Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback, arXiv, 2204.05862 (2022)
- (4) Rafailov, R., et al. : Direct Preference Optimization: Your Language Model is Secretly a Reward Model, Advances in Neural Information Processing Systems, **36**, 53728~53741 (2023)
- (5) Schulman, J., et al. : Proximal Policy Optimization Algorithms, arXiv, 1707.06347(2017)
- (6) Hugging Face : LLM-jp
<https://huggingface.co/llm-jp>
- (7) Touvron, H., et al. : Llama 2: Open Foundation and Fine-Tuned Chat Models, arXiv, 2307.09288 (2023)
- (8) Zheng, L., et al. : Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena, Advances in Neural Information Processing Systems, **36**, 46595~46623 (2023)
- (9) Chen, T., et al. : TVM: An Automated End-to-End Optimizing Compiler for Deep Learning, 13th USENIX Symposium on Operating Systems Design and Implementation, 578~594 (2018)
- (10) PyTorch : Announcing the 2024 PyTorch Contributor Awards
<https://pytorch.org/ecosystem/contributor-awards-2024>
- (11) GitHub : ggml-org/llama.cpp
<https://github.com/ggml-org/llama.cpp>



LLMを用いたラダープログラムの処理内容説明の生成技術

LLM-Based Program Analysis for Ladder Logic

*情報技術総合研究所

要旨

ラダー言語はPLC(Programmable Logic Controller)で制御する内容を記述するプログラミング言語であり、様々な設備の制御に使用されている。制御設計者が設備を改修する際にはラダー言語で記述された制御プログラム(以下“ラダープログラム”という。)を解読して処理内容を理解する必要があるが、この作業は負担が大きい。

そこで三菱電機は、LLM(Large Language Models)を用いてラダープログラムを解析し、処理内容の説明を生成する技術を開発した。ラダープログラムだけでなくPLCのログデータも合わせて解析し、信号の変化順にラダープログラムの処理内容を説明することが特長である。これによって設備内の機器の制御順序が分かりやすい説明になり、設計者のラダープログラム解読の負担を軽減できる。

1. まえがき

PLCは様々な設備の制御に使用されている。PLCが設備の制御を行う際には、設備内の機器から信号を受け取って、信号を制御プログラムで処理し、設備内の機器に対して信号を出力する、ということを繰り返す。制御プログラムを記述するプログラミング言語は幾つか存在するが、日本ではラダー言語というプログラミング言語が特に多く使用されている。

図1を用いて、ラダープログラムによる設備制御の概要を述べる。図1(a)は、PLCと機器(スタートスイッチ、プッシャー)で構成された設備を表しており、図中のX0、Y10などはPLCと機器の間でやり取りされる入出力信号である。このようなPLCと機器で構成された設備では、PLCは機器から入力信号を受け取って、その信号をラダープログラムで処理して出力信号を機器に渡すことで、機器の制御を行う。

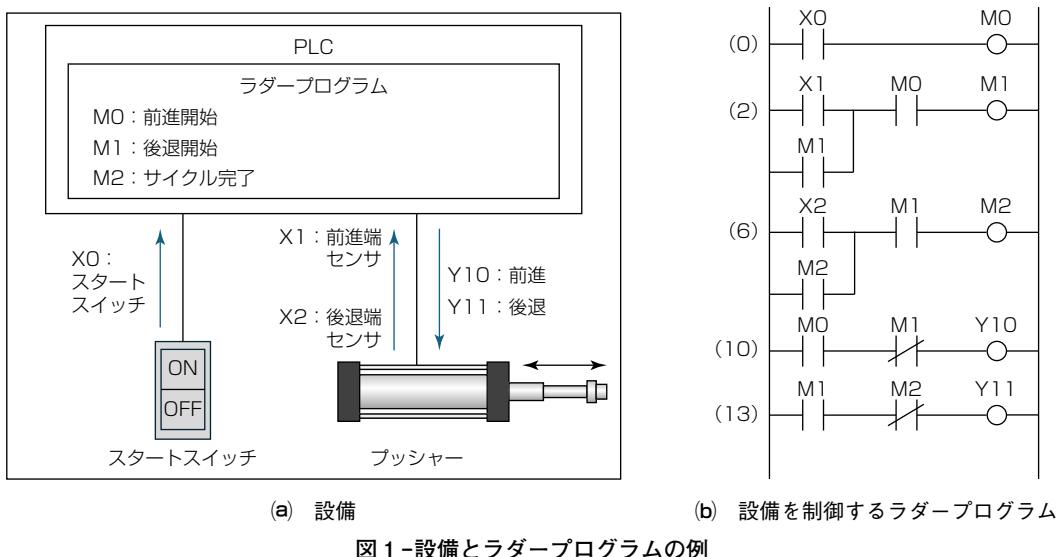


図1-設備とラダープログラムの例

図1(b)は、図1(a)の設備を制御するラダープログラムの例である。ラダープログラムでは左右の縦線の間に横線を引いて、横線の左側に条件、右側に処理を記述して、どのような場合にどのような処理を行うかを表す。例えば、図1(b)の(0)について、左の部分は“X0が1のとき”という条件、右の部分は“M0を1にする”という処理を示している。同様に、図1(b)の(10)では、左の部分で“M0が1、かつM1が0のとき”という条件、右の部分は“Y10を1にする”という処理を示して

いる。図1(b)の(0), (10)の条件、処理を組み合わせると“X0が1、かつM1が0のときY10を1にする”という制御になる。この制御によって、設備のスタートスイッチをONにすると、プッシャーは前進端に到達するまで前進する。

このように、ラダープログラムは複数の入力信号の値やその変化順序が処理に影響し、かつ一つの出力信号に関係する処理がラダープログラムの複数箇所に分けて書かれているため、解読が難しい。

2. ラダープログラムの制御内容理解の課題

制御設計者が設備を改修する場合や、設備の不具合の原因を推定する場合には、設備がラダープログラムによってどのように制御されているのかを理解している必要がある。しかし、長く使われている設備では、当時設計を行った制御設計者が退職などで在籍していない場合があり、ラダープログラムの解読が必要になる。

図2は、図1(b)のラダープログラムを人が解読した結果を文章で表した例である。この文章では、装置の信号がどのような順序で変化するか、信号変化がラダープログラムのどの部分によって発生するかが分かるようになっている。文中のステップは図1(b)の左に書かれた番号と対応している。このような処理内容の説明があれば、制御設計者が設備の改修を行う場合や、設備の不具合の原因を推定する場合に、ラダープログラムのどの部分を確認、修正すればよいか判断できる。

しかし、1章で述べたように、ラダープログラムの解読は難しく、制御設計者が大規模なラダープログラムを解読して図2のような説明を作る作業は負担が大きい。

そこで当社では、LLMを用いてラダープログラムの処理内容説明を生成する技術を開発した。この技術で生成したラダープログラムの処理内容説明は、設備内の機器をどのような順序で制御しているのか分かりやすいことが特長である。

ユーザーがスタートスイッチをONにするとX0が1に変化し、プログラムの0ステップ目の処理によりM0(前進開始)が1に変化する。M0(前進開始)が1に変化するとプログラムの10ステップ目の処理によりY10(前進)が1に変化し、プッシャーが前進を開始する。

プッシャーが前進端に到達するとX1(前進端センサ)が1に変化し、プログラムの2ステップ目の処理によりM1(後退開始)が1に変化する。M1(後退開始)が1に変化するとプログラムの13ステップ目の処理によりY11(後退)が1に変化し、プッシャーが後退を開始する。また、プログラムの10ステップ目の処理によりY10(前進)が0に変化するため、プッシャーの前進は停止する。

プッシャーが後退端に到達するとX2(後退端センサ)が1に変化し、プログラムの6ステップ目の処理によりM2(サイクル完了)が1に変化する。M2(サイクル完了)が1に変化するとプログラムの13ステップ目の処理によりY11(後退)が0に変化し、プッシャーの後退が停止する。

ユーザーがスタートスイッチをOFFにするとX0が0に変化し、プログラムの0ステップ目の処理によりM0(前進開始)が0に変化する。M0(前進開始)が0に変化するとプログラムの2ステップ目の処理によりM1(後退開始)が0に変化する。M1(後退開始)が0に変化するとプログラムの6ステップ目の処理によりM2(サイクル完了)が0に変化する。

図2-ラダープログラムの処理内容の人による解説結果例

3. ラダープログラムの処理内容説明の生成方法

LLMを用いてプログラムの説明を生成する場合、基本的にラダープログラムの上から順に説明が生成される。そのため、ラダープログラムだけを入力データとしてLLMで処理内容説明を生成するだけでは、図2に示したような、ラダープログラムで行われる処理の順序が分かる説明が生成できない。この技術ではラダープログラムだけでなくPLCのログデータを入力データとして用いることで、この課題を解決する。ラダープログラムの処理内容説明を生成するまでの手順と、各手順での入力データ、出力データを、表1に示す。

表1-この技術での処理内容説明の生成手順

No.	手順	入力データ	出力データ	説明
1	ラダープログラムの分割	ラダープログラム	分割したラダープログラム	3.1節参照
2	信号変化順の抽出	ログデータ	各時刻での信号変化の情報	3.2節参照
3	信号変化とラダープログラムの対応付け	No.1とNo.2の出力データ	信号変化とラダープログラムの関係情報	3.3節参照
4	処理内容説明の生成	No.3の出力データと、各信号の意味	ラダープログラムの処理内容説明	3.4節参照

3.1 ラダープログラムの分割

図2の解読結果では、最初にラダープログラムの0ステップ目の処理について説明し、その次にラダープログラムの10ステップ目の処理が行われることを説明している。ラダープログラムでは条件を満たしている部分だけ処理が行われるため、このようにプログラムの処理の並び順と実際に処理が実行される順序が一致しない場合がある。

そこで、この技術では並び替えを可能にするために、図3のように、ラダープログラムの分割を行う。分割は、処理の内容と、その処理を行う条件の組合せである回路ブロック単位で行う。そして、後の手順で正しい順番での並び替えを可能にするために、条件部分と処理部分についても分割する。

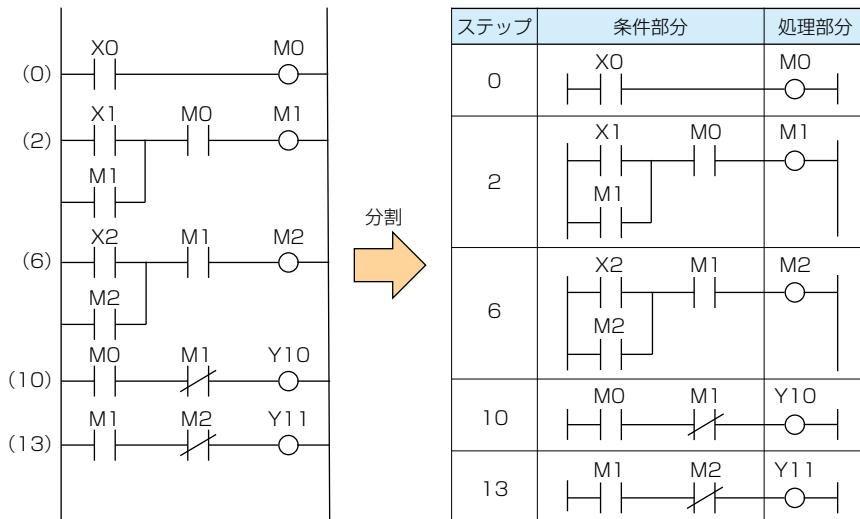


図3-ラダープログラムの分割イメージ

3.2 信号変化順の抽出

図2の解読結果では、最初にX0が1になり、その後X1が1になるなど、入力信号の変化順が示されている。このような入力信号の変化は機器の動きによって起こるものであり、ラダープログラムからは入力信号がいつどのように変化するかを読み取ることができない。

そこで、この技術では、ラダープログラムに加えて、PLCのログデータを入力データとして使用する。PLCのログデータは設備動作中の各信号の状態を時刻とともに記録したもので、図4のように、表形式で表すことができる。例えば図4では、時刻列が00:01.0の行でX0列の値が1であることから、00:01.0の時刻でX0の値が1であることが読み取れる。

この技術では、図4のように、各信号の値が前の時刻から変化している部分をログデータから抽出することで、処理内容の説明の生成に必要になる、信号がいつどのように変化したのかを表す信号変化順の情報を作成する。

前時刻から値が変化している部分								
時刻	X0	X1	X2	Y10	Y11	M0	M1	M2
00:00.0	0	0	1	0	0	0	0	0
00:01.0	1	0	1	1	0	1	0	0
00:02.0	1	0	0	1	0	1	0	0
00:03.0	1	0	0	1	0	1	0	0
00:04.0	1	1	0	0	1	1	1	0
00:05.0	1	0	0	1	1	1	1	0
00:06.0	1	0	1	0	0	1	1	1
00:07.0	0	0	1	0	0	0	0	0

ログデータ

時刻	変化
00:01.0	X0が1に変化 Y10が1に変化 M0が1に変化
00:02.0	X2が0に変化
00:04.0	X1が1に変化 Y10が0に変化 Y11が1に変化 M1が1に変化
00:05.0	X1が0に変化
00:06.0	X2が1に変化 Y11が0に変化 M2が1に変化
00:07.0	X0が0に変化 M0が0に変化 M1が0に変化 M2が0に変化

信号変化順

図4-ログデータからの変化抽出イメージ

3.3 信号変化とラダープログラムの対応付け

図2の解読結果では、信号変化とラダープログラムの該当部分を対応付けて説明している。このような説明を生成するためには、3.1節の処理でラダープログラムの分割を行った結果と、3.2節の処理で信号変化的抽出を行った結果の対応付けを行う必要がある。

この技術では、対応付けの候補を抽出し、候補になった対応付けが正しいかをLLMを用いた検証によって確認する。

対応付けの候補を見つける手順では、3.2節の処理で抽出した信号変化を基に3.1節の処理で分割したラダープログラムから、条件部分の信号と、処理部分の信号の両方が同時刻で変化しているプログラムを抽出する。図5の例では、00:01.0の時刻では三つの信号変化が発生しているが、条件部分と処理部分の両方にこれらの変化が含まれるラダープログラムは、ステップ0とステップ10のラダープログラムである。この二つのプログラムで、条件となる信号の値が変化することで処理が行われ、他の信号の値が変化した可能性があると予測できる。

候補になった対応付けが正しいかをLLMを用いて検証する手順では、信号の値とラダープログラムをLLMに与えて、ラダープログラムの処理で信号の値が変化するかを確認する。この手順が必要である理由は、先に述べた、抽出した対応付けの候補が正しいとは限らないためである。例えば、図5のステップ10のプログラムは条件部分にM1の信号が含まれており、M1の信号の値次第では、M0が1になってもY10が1にならない。また、同じ信号の値を変化させる処理の候補がラダープログラム内に複数存在する場合には、実際に信号を変化させた部分がどの候補かを、この方法で確認する必要がある。

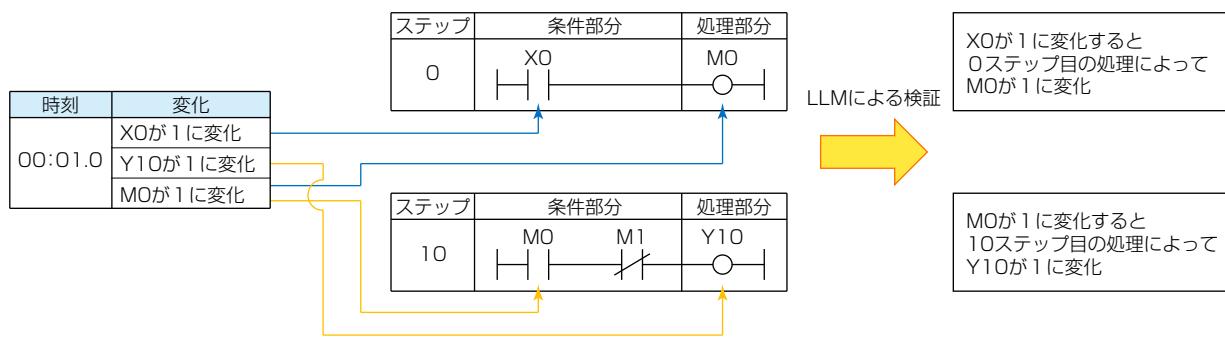


図5-信号変化とラダープログラムの対応付けイメージ

3.4 処理内容説明の生成

最後に、3.3節で得られた説明を結合して、ラダープログラムの処理内容説明を生成する。このとき、“X0(スタートスイッチ)”のように、信号だけでなく各信号の意味も説明に加えることで、設備の動作との対応を分かりやすくしている。

4. 検 証

この技術によって、ラダープログラムの処理内容を正しく記述した説明が生成できること、及び信号の変化順に処理内容を説明した分かりやすい説明が生成できることを確認するために、検証を行った。検証ではLLMとしてgpt-4o(2024-05-13)を使用した。

4.1 処理内容を正しく記述した説明が生成できることの検証

この技術で図1(b)のラダープログラムの処理内容説明を生成し、その結果を図2の人による解読結果と比較することで、処理内容を正しく記述した説明が生成できているかを検証した。図6は、この技術によるラダープログラムの処理内容説明の生成結果である。図2と図6を比較すると、“プッシャーが前進端に到達”のような信号変化の原因は入力データに情報がないため図6では記載されないが、“前進”“前進端センサ”的な各信号の意味からおおよその動作の予測は可能である。それ以外は図2と図6で同等の内容が記述されており、処理内容を正しく記述した説明が生成されている。

X0(スタートスイッチ)が1に変化すると、プログラムの0ステップ目の処理によりM0(前進開始)が1に変化します。
M0(前進開始)が1に変化すると、プログラムの10ステップ目の処理によりY10(前進)が1に変化します。

3.0秒後X1(前進端センサ)が1に変化すると、プログラムの2ステップ目の処理によりM1(後退開始)が1に変化します。
M1(後退開始)が1に変化すると、プログラムの13ステップ目の処理によりY11(後退)が1に変化します。
さらにプログラムの10ステップ目の処理によりY10(前進)が0に変化します。

2.0秒後X2(後退端センサ)が1に変化すると、プログラムの6ステップ目の処理によりM2(サイクル完了)が1に変化します。
M2(サイクル完了)が1に変化すると、プログラムの13ステップ目の処理によりY11(後退)が0に変化します。

1.0秒後X0(スタートスイッチ)が0に変化すると、プログラムの0ステップ目の処理によりM0(前進開始)が0に変化します。
M0(前進開始)が0に変化すると、プログラムの2ステップ目の処理によりM1(後退開始)が0に変化します。
M1(後退開始)が0に変化すると、プログラムの6ステップ目の処理によりM2(サイクル完了)が0に変化します。

図6-この技術によるラダープログラム処理内容説明の生成結果

4.2 信号の変化順に処理内容の説明を生成できることの検証

この技術を用いて生成した処理内容説明と、ラダープログラムだけを入力データとして生成した処理内容説明を比較し、3章で述べた手順の効果を確認した。ラダープログラムだけを入力データとして生成した処理内容説明を図7に示す。

図7に示した説明はラダープログラムの上から順に説明を作成しているため、プログラムのどの部分がどの順序で処理を行うかが判断できない。また、各信号の変化がどの順序で発生するかも分からず。一方、図6に示すように、この技術を用いて生成した処理内容説明ではこれらの問題を解決し、ログデータでの信号の変化順にラダープログラムの処理内容を説明できている。このような説明があれば、ラダープログラムの解読にかかる設計者の負担を軽減可能である。

1. X0がONのとき、M0をONにします。
2. X1がON、またはM1がON、かつM0がONのとき、M1をONにします。
3. X2がON、またはM2がON、かつM1がONのとき、M2をONにします。
4. M0がON、かつM1がOFFのとき、Y10をONにします。
5. M1がON、かつM2がOFFのとき、Y11をONにします。

図7-ラダープログラムだけを入力データとした処理内容説明の生成結果

5. むすび

LLMを活用し、ログデータから抽出した信号変化とラダープログラムを対応付けることによって、設備内の機器の制御順序が分かりやすい説明を生成する技術について述べた。この技術を活用することによって、設計者のラダープログラム解読の負担を軽減できる。

今後は処理内容についての質問に回答するチャットボットや仕様との差異の説明機能など、この技術を更に発展させた技術についても検討する。

AI活用によるLSI開発の高度化

AI-driven Enhancement in LSI Development Technology

森 敦弘* 深井哲人*

Atsuhiro Mori Yoshito Fukai

高品信秀*

Nobuhide Takashina

南口比呂*

Hiro Minamiguchi

*設計技術開発センター

要 旨

三菱電機では、製品の高性能化・高機能化の要になるLSI(Large Scale Integration: 大規模集積回路)を社内開発している。近年、LSI製造プロセスの微細化やチップレット技術の進化によって、LSIに内蔵される回路は急激に大規模化の一途を辿(たど)り、増大する開発コストの削減が急務である。LSI開発はこれまで自動化による開発効率化が進められてきたが、依然として熟練技術者の経験やノウハウが必要な工程が残っている。そこで熟練技術者への依存を解消するため、自然言語処理AIを活用することで、社内の膨大な設計資産を効率的に活用できる設計支援システムの構築に取り組んでいる。また、LSI開発に生成AIを活用できる技術者も育成している。

1. まえがき

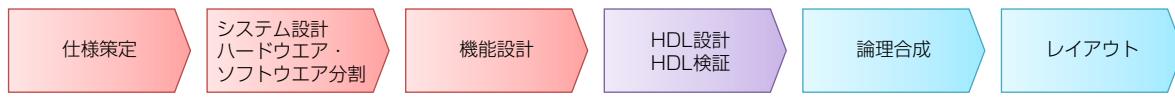
組み込みAIや5G(第5世代移動通信システム)等の新規アプリケーションでは、LSIの大規模化が進んでいる。LSI開発は自動化が比較的進んでいるものの、熟練技術者に依存する部分も多く、労働人口の減少に伴いノウハウの伝承も困難になりつつある。そこで当社では発展が著しいAI技術を活用したノウハウの形式知化に取り組んでいる。特に製品固有の知見が必要なシステム設計等の分野で独自のAI活用手法を開発することで、設計者を支援するシステムを整備してきた。

一方、2022年のChatGPT^(注1)リリースをきっかけに、生成AI技術を容易に応用できるようになった。これによって、AIは人の作業を支援するだけでなく、設計成果物を提供するものへと大きく進化した。生成AIは、文章や絵を自動生成するだけでなく、LSI回路も設計できる可能性があり、期待が高まっている。そこで当社では、LSI開発への生成AIの活用と技術者の育成に取り組んでおり、本稿ではこれらの事例について述べる。

(注1) ChatGPTは、OpenAI OpCo, LLCの登録商標である。

2. LSI開発の課題

図1にLSI開発の工程概要を示す。まず、LSIを使用する製品システムの要求仕様を分析し、LSIの仕様策定やシステム設計を経て、ASIC(Application Specific Integrated Circuit)やFPGA(Field-Programmable Gate Array)によるハードウェア処理とCPU等によるソフトウェア処理に機能を分割する。次に分割した各々について詳細に機能設計し、HDL(Hardware Description Language)を用いてLSIの回路を設計して検証する。この後、HDL記述から論理合成と呼ばれる工程でゲートレベル回路に変換し、シリコン上へ転写する回路パターンをレイアウトすることで、LSIの設計データが完成する。なお、これらの工程のうち、回路シミュレーターを用いて検証する工程や、HDL記述から論理合成する工程では設計の自動化が進んでいる。しかしながら、LSI開発工程全体を見渡すと、次に示す課題が残っている。



2.1 熟練技術者に依存した開発

仕様策定やシステム設計といった上流工程では、製品の要求仕様を分析する必要があり、製品開発の過去事例に精通した熟練技術者に依存していた。例えば、要求仕様からのハードウェア処理とソフトウェア処理の分割検討(図2)には、複

数の信号処理システムの開発経験や、ハードウェア処理及びソフトウェア処理全般の幅広い知見を持つ、熟練技術者の参画が必須である。

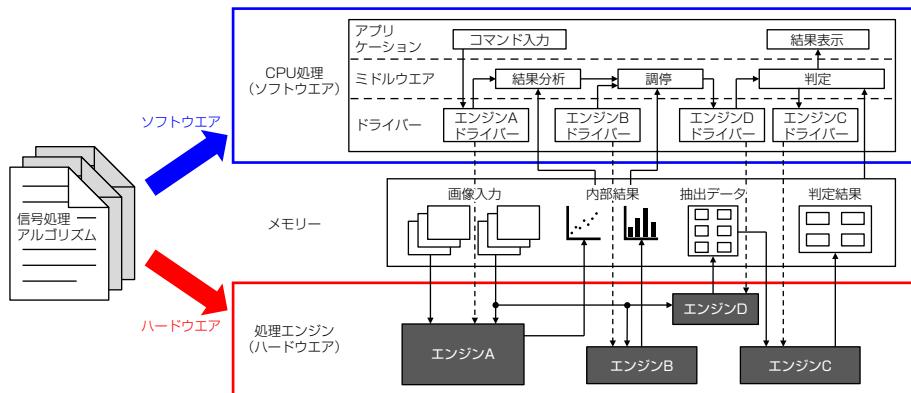


図2-LSI開発でのハードウェア処理とソフトウェア処理分割の検討事例

2.2 過去設計のノウハウ活用

これまでのLSI開発では、設計の効率化や不具合防止を目的とした、過去設計で得たノウハウの活用が不十分であった。例えば、過去の開発案件でのDR(Design Review)の指摘事項やFMEA(Failure Mode and Effect Analysis)の結果といった、ノウハウ資料を設計者が手で探していたため、確認項目の抽出漏れが発生し、検証が不十分になるおそれがあった。抽出漏れを防止するには、確認項目の自動抽出が必要である。

3. AIで設計者を支援するシステム

2章に述べた課題を解決するため、当社では社内のAIソリューションプラットフォーム⁽¹⁾を活用し、設計者を支援する独自システムを、次のような方針で構築している。まず、過去の開発データ(仕様書、設計書、不具合事例等)をAIに学習させて、推論モデルを構築する。次に推論モデルを用いたシステムに対して、設計者が新規に設計する回路の機能や仕様などの情報を入力することで、性能や品質に影響を与えるドキュメントやキーワード、最適パラメーターなどを寄与度の高い順に出力する。このように、今回開発するシステムは、従来、熟練技術者の暗黙知になっていたノウハウを活用することで、設計者が製品固有の仕様に最適な設計や検証ができるようになるものである。システムの概念図を図3に示す。これらの方針の下、LSI開発の各課題を解決するため開発を進めている事例について次に述べる。

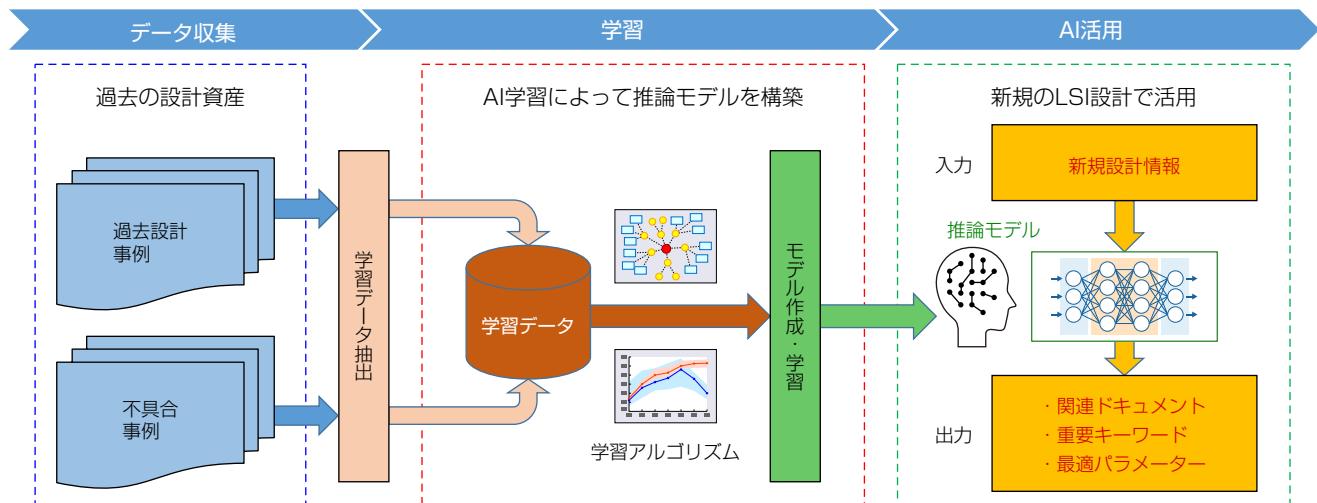


図3-設計者をAIで支援するシステムに関する概念図

3.1 システム事例1：過去設計資産検索システム

現在開発を進めている設計支援システムの一つ目は、過去設計資産検索システムであり、新たなLSI開発の仕様策定やシステム設計、及びハードウェア処理とソフトウェア処理の分割検討等に必要な情報収集の効率化を実現する。具体的には、事前に過去の開発での仕様書や内部設計書、検証仕様書などをシステムに学習させておき、設計対象の新規回路に関連したキーワードを設計者がシステムに入力すると、そのキーワードや類義語を含む設計資料を検索し、提示するシステムである。システム画面を図4に示す。さらに、このシステムでは設計者が入力した複数のキーワード間のつながりを示す相関図を作成する。このシステムを用いることで、製品開発の過去事例に精通した熟練技術者に頼ることなく、設計対象の回路仕様やソフトウェアによるLSIの制御仕様、またLSI上の他の回路ブロックとのデータ授受や連携動作等といった幅広い知見を短時間で得ることが可能になる。

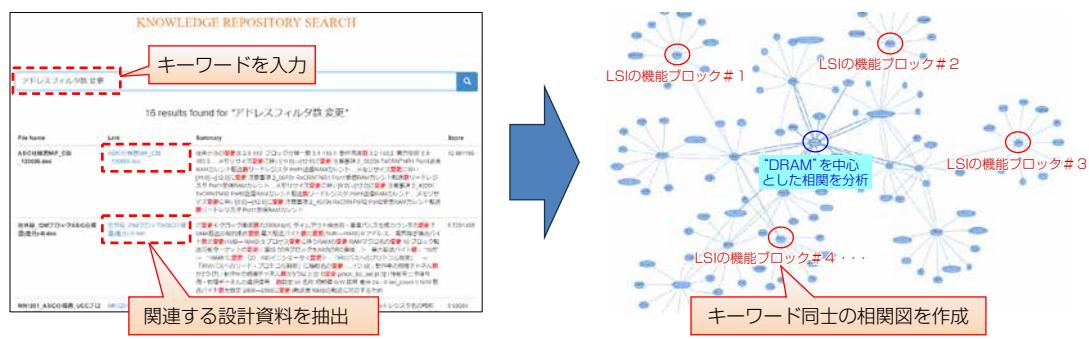


図4-現在開発中のシステム事例：過去設計資産検索システム

3.2 システム事例2：過去不具合検索システム

現在開発を進めている設計支援システムの二つ目は、過去不具合検索システムである。このシステムは、3.1節に述べた過去設計資産検索システムと同様に、過去のノウハウ資料である不具合事例集や関連する設計ドキュメント類を自然言語処理AIに学習させたものである。具体的には、事前に過去の不具合に関する説明と対応する解析結果・対処方法などを学習させておき、設計対象の新規回路に関連したキーワードや文章を設計者がシステムに入力することで、関連する過去不具合事例を検索して提示することによって、設計品質を向上させるシステムである。システム画面を図5に示す。さ

図5-現在開発中のシステム事例：過去不具合検索システム

らに、このシステムでは、抽出した不具合の原因や影響度、不具合発見工程などを円グラフ上の複数のカテゴリーに分類して可視化する。例えば、不具合の原因であれば、仕様検討の誤り、設計の誤り、HDL記述の誤りなどのカテゴリーに分類することで、不具合発生の条件や傾向等が分かり、設計者に注意を促す。このシステムを用いることによって、過去事例の活用を促進し、設計者が人手で過去事例を確認する場合に比べて、確認漏れの防止が可能になる。さらに、DRやFMEA等の開発ホールドポイントで過去事例に基づく指摘が可能になるため、設計手戻りの抑制にもつながる。なお、AI活用によるソフトウェア開発向け類似不具合検索に関する取組みは、本号“生成AIを活用した社内技術文書検索システムの構築と展開”の3章に述べている。

4. LSI開発への生成AI活用

LSIの機能設計では、HDLでソースコードを記述する。HDLはプログラミング言語をベースに開発したものであり、回路図に比べて抽象度が高い。ソースコードの記述では、昨今、ChatGPTなどの大規模言語モデル(生成AI)を用いた指示(プロンプティング)によってソースコードを自動生成する事例や、生成AIのプラグインを持つソースコードエディターのAIアシスタント機能を活用してソースコード記述案を設計者に提示する事例などが報告され始めている。そこで、ソースコードの記述に生成AIを活用した事例を参考に、HDL開発への生成AI活用にも取り組んでいる。

4.1 生成AI活用の具体的な取組み

具体的な取組み内容を図6に示す。次の(1)～(3)に示す三つを生成対象として各手法を試行した。

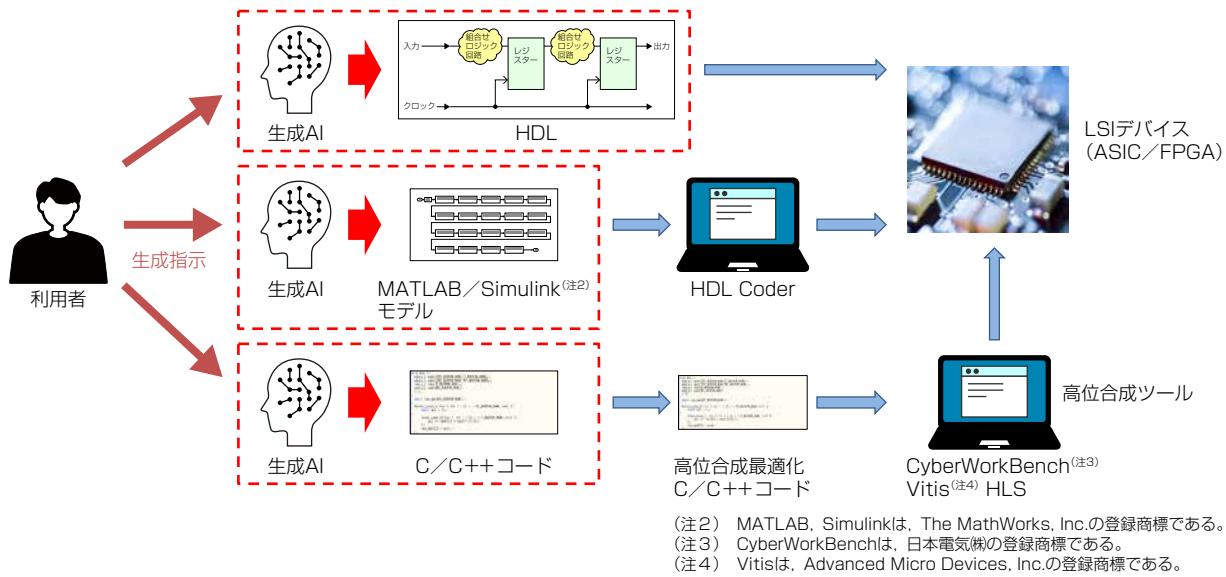


図6-HDL設計及びHDL検証への生成AI適用概念図

(1) HDLソースコードの生成

設計者が回路の機能仕様などを生成AIにプロンプティングし、HDL記述を直接生成する手法。

(2) MATLABソースコードの生成

モデルベース開発ツール：MATLAB/SimulinkのHDL Coderを用いてHDL記述を得る設計フローで、設計者が機能仕様などを生成AIにプロンプティングし、MATLABソースコードを生成する手法。

(3) 高位合成可能なC/C++ソースコードの生成

C/C++ソースコードから高位合成ツールを用いてHDL記述を得る設計フローで、設計者が機能仕様などを生成AIにプロンプティングし、高位合成可能なC/C++ソースコード及び最適化用プラグマを生成する手法。

これら三つのどの手法でも、有用性を確認できた。しかしながら、現状ではHDL記述を得る際に、HDL Coderや高位合成ツールで実行エラーが発生するソースコードを生成したり、人手による設計と比較して回路の性能が悪化したりするなどの課題も多い。今後の更なる調整や、生成AIの進化を取り込むなど、継続的な改善が必要である。なお、生成AI活

用によるLSI開発に関する他の取組みは、本号“生成AIを活用した社内技術文書検索システムの構築と展開”の4章に述べている。

4.2 生成AIをLSI開発に適用できる技術者の育成

当社では要素技術開発部門の横断的な取組みとして、ソフトウェアやLSIのソースコードを開発する技術者に、プリント基板設計や機械設計の技術者も加えて、最新の生成AI技術とその実践的な活用方法について、社内で活用しているプログラムや設計事例を題材にワーキンググループや勉強会⁽¹⁾を開催し、知見を深めている。このような取組みを通じて、設計技術開発センターでは、生成AIをLSI回路設計や検証に適用できる専門人材を育成している。今後、専門人材がワーキンググループや勉強会等で得たLSI開発事例を教材として全社の各事業部門に展開することで、各事業部門のLSI技術者の育成を推進していく方針である。また、当社のデジタル基盤“Serendie”⁽²⁾が提供する生成AIフレームワークを活用したLSI開発向け設計支援ツールも開発しており、各事業部門に生成AIを活用したLSI開発手法の浸透を図っていく。

5. むすび

当社でのAIを活用したLSI開発高度化の試行事例として、AIで設計者を支援するシステムの構築やLSI開発への生成AI適用などについて述べた。今後は、製品に実装するLSI開発への適用を増やしていく中で設計者からの改善要望を集めて、より実用性の高いシステムに仕上げていく。また、AI技術は今後ますますの発展が予想され、より複雑なLSI開発工程への適用も可能になると考えられる。当社では今後も最新の技術を取り入れて、LSI開発の更なる高度化に取り組んでいく。

参考文献

- (1) 玉谷基亮：三菱電機グループの持続的なものづくりを支えるAIソリューション群、三菱電機技報、98、No.9、3-01～3-05 (2024)
- (2) 三菱電機：Serendie セレンディ
<https://www.mitsubishi-electric.co.jp/serendie/>

特集論文

製造プロセスへのAI適用

AI Application to Manufacturing Processes

氏家達浩*

Tatsuhiro Ujiiie

横堀祐芽*

Yuka Yokobori

高阪 翔*

Kakeru Kosaka

榎原正朗*

Masaaki Sakakibara

*生産技術センター

要 旨

生産現場でのAI運用は、生産性向上と品質安定化に大きな社会的意義を持つ。世界の製造業では、データ駆動型の意思決定やプロセス最適化にAIを導入する取組みが加速しているが、実際の生産ラインへの実装には課題が残る。

三菱電機は、加工・組立て・検査などの工場ライン作業への高効率AI運用技術を開発している。具体的には、樹脂成形の条件導出技術、機械加工パラメーターの自動最適化、溶接パラメーターの自動最適化など、製造プロセスの各段階でAIを適用することで、生産効率と品質の向上に取り組んでいる。これらの技術によって、製造業のデジタル変革を加速させて、持続可能な生産体制の構築に貢献する。

1. まえがき

AIは世界中で急速に普及しているが、製造業の現場では他分野より限定的で、多くの企業ではPoC(Proof of Concept)にとどまっている。

当社では熟練者の知識や経験を分析し、明示的なルールを数値データとして捉え直した上で、AIに学習させる技術開発に注力している(図1)。



図1-生産現場でのAI実用化

加工プロセスへの適用では、①樹脂成形の条件導出技術、②機械加工パラメーターの自動最適化、③溶接パラメーターの自動最適化の開発を進めてきた。これによって、属人的技能依存から脱却し、品質の標準化・生産の安定化・技能伝承問題の解消に貢献している。

本稿では、当社の製造プロセスへのAI適用事例を述べて、生産現場でのAI実用化の展望について述べる。

2. 樹脂成形の条件導出技術

樹脂成形は、製品の外観品質と機能性を左右する重要技術である。高精度な筐体(きょうたい)や精密部品の製造を可能にして、コスト競争力と製品差別化を同時に実現する。また、多様な製品群の迅速な市場投入を支える基盤技術として不可欠である。

2.1 樹脂成形パラメーター調整の課題

樹脂成形では、条件設定が熟練者の経験と勘に依存し、エンジニアの技術水準が品質とリードタイムに影響する。当社でもトライアンドエラーによる条件調整が、リードタイム長期化と品質ばらつきの原因になっている。

2.2 ベイズ最適化を用いた樹脂成形条件導出技術

この課題解決のため、未知関数にも適用可能なベイズ最適化を活用した樹脂成形条件導出技術を開発した。成形品質の定量化が重要だが、特に外観品質は目視検査が一般的で、判断基準が不明瞭のため作業者間のばらつきが課題であった。

この解決のため、成形品質の定量化に取り組んだ。一つ目は樹脂収縮による凹み(くぼみ)(ヒケ)で、型内センサーの値と表面粗さ計による凹み量との相関からヒケ特微量という指標を見いだした。二つ目は流動の乱れによる光沢不良(フローマーク)で、画像処理技術を用いて熟練者の官能評価と相關する指標を得た。これによって、図2のように成形加工とセンシング、品質評価、AIによる条件候補導出を繰り返すことで、機械的に最適成形条件を導出することが可能になった。

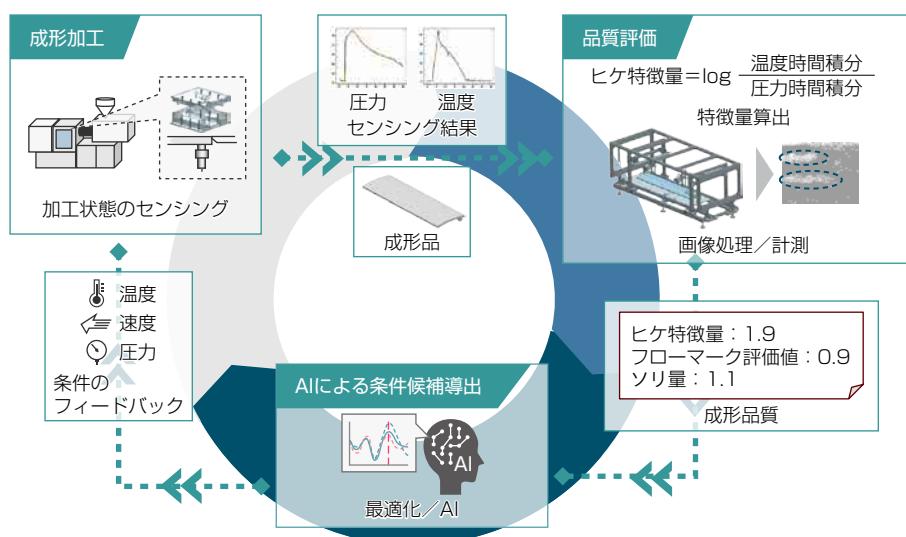


図2-ベイズ最適化を用いた樹脂成形条件導出技術の概略図

2.3 ベイズ最適化を用いた樹脂成形条件導出技術の適用例

この技術の検証内容を図3で示す。長尺板状意匠部品を対象として、成形条件の導出能力をこのシステム、熟練者、非熟練者で比較した。対象とする不良はソリ、ヒケ、フローマークの三つとして、それらが要求品質値を満たせば良品と判定した。

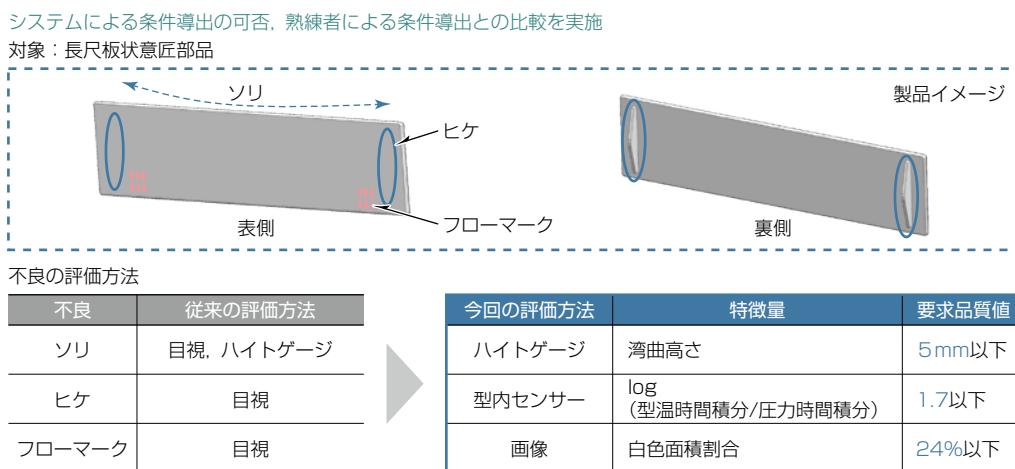


図3-検証方法

比較結果を図4で示す。それぞれ導出した成形条件に基づいた成形品について、評価結果をまとめた。また、必要になった変更回数も示す。その結果、システムと熟練者は要求品質を満たす成形条件を導出できたが、非熟練者は導出できなかった。さらに、システムは熟練者をわずかに上回る変更回数で同等の品質を達成できた。



図4-AI適用による効果確認

2.4 今後の展望

この研究で開発した樹脂成形の条件導出技術は、属人的技能依存からの脱却に向けた第一歩として一定の成果を上げた。今後は、AI判断根拠の可視化を更に強化し、熟練エンジニアの判断プロセスを再現できるようにして、エンジニア育成の加速を図る。これらの取組みによって、熟練者の経験や勘に依存しない科学的アプローチに基づいた樹脂成形技術を確立し、製造業全体の技術基盤強化と競争力向上に貢献していく。

3. 機械加工パラメーターの自動最適化

機械加工は製造業の根幹を支える基盤技術であり、製品の寸法精度や表面品質を決定する。企業競争力の源泉になる高精度部品の安定生産、コスト削減、及び短納期化を実現するためには、機械加工パラメーターの最適化が不可欠である。

3.1 機械加工パラメーター調整の課題

機械加工では、図面で要求される寸法公差、幾何公差、表面粗さを満たす加工条件設定が重要である。新製品形状の加工では、熟練者でもワーク形状などによって精度が変わるために、複数回の試加工が必要になる。

このような試行錯誤は材料コスト増加、設備稼働率低下、リードタイム長期化を招く。さらに近年の多品種少量生産化や製品ライフサイクル短縮化によって、新規加工条件設定の頻度が増加し、迅速かつ適切な条件設定が課題になっている。

3.2 機械加工パラメーターの自動最適化手法

この課題解決のため、加工条件から加工精度を予測する自動最適化システムを開発した(図5)。このシステムはマシニングセンターのエンドミル加工を対象として、加工条件と加工精度をデータベース化して機械学習によって関係性をモデル化している。入力変数にはワークの形状情報(厚み、高さなど)と加工条件(送り速度／軸切り込みなど)を使用した。各品質指標の特徴量を求めて、ワーク形状及び加工条件の影響度を明らかにした。例えば平面度では、壁高さや壁厚みの形状情報が重要な特徴量になった。モデル構築では直交表や機械学習特徴選択による要素絞り込みを実施し、データを効率的に活用した。

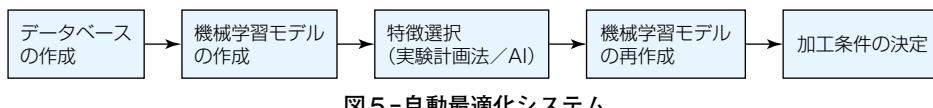


図5-自動最適化システム

3.3 自動最適化の適用例

開発システムを、簡易化した製品模擬ワークの実加工に適用した(図6、図7)。表面粗さの予測精度は約30%以内に収まり、軽切削領域を除いて、ワーク形状や加工条件変更時の寸法・幾何公差の増減傾向が実測と予測で合致した。この予測モデルによって加工前に加工条件と精度の関係把握が可能になった。このシステムの利用で、従来の経験ベース手法(約380min/条件)と比較し、機械学習を用いたこのシステムでは192min/条件になり、約49%の時間削減が実現できる。



図6-製品模擬ワーク



図7-製品模擬ワークのエンドミル加工

3.4 今後の展望

この取組みでは一定の成果を得たが、更なる高度化を目指している。具体的には、加工状態変化のセンシング手法を確立し、微小変化の影響を学習モデルへ反映する。これによって、寸法・幾何公差の予測精度向上と要求品質を満たす最速加工条件の導出を目指す。この技術は多品種少量・高精度が求められる大型ワーク製造に有効である。加工前の品質・能率評価は製造現場のニーズに合致しており、加工条件最適化によって高品質製品の安定供給とコスト競争力強化を両立していく。

4. 溶接パラメーターの自動最適化

溶接は製造業の主要接合技術であり、多くの製品で不可欠なプロセスである。溶接部の品質は製品の信頼性に直結するため、適切なパラメーター設定は安全性と品質向上に重要である。また、溶接条件の自動最適化によって、製造コスト削減と生産性向上が実現できる。

4.1 溶接パラメーター調整の課題

溶接プロセスでは、電流・電圧・速度などの多様なパラメーターが品質や形状精度に影響する。従来、これらの調整は熟練技術者の経験に依存し、最適条件の導出には多くの試作が必要であった。市販シミュレーションも存在するが、実機との乖離(かいり)があり実用的な予測が難しいため、製造現場でのパラメーター最適化が課題になっている。

4.2 自動最適化手法

当社では高精度独自データベースとAIモデルを組み合わせた溶接パラメーター自動最適化システムを開発した。核になるのは熱弾塑性解析の代理モデルで、溶接歪(ひず)みを推定するニューラルネットワークを構築した(図8)。溶接条件、材料特性、継手形状を入力として、変形量を出力として学習させている。標準データベース使用時の実機との乖離には、実測データを織り込んだ独自データベースで対応した。このシステムによって、溶接パラメーターの最適化を効率的に実施でき、実験と計測に依存せず開発プロセスの高速化と高精度化を同時に達成している。

4.3 自動最適化適用例

遮断器筐体モデルのケーススタディーによって、標準データベースを用いた従来手法と、歪み実績値を織り込んだ独自データベースの解析結果を比較した(図9)。独自データベースは実機の変形挙動に近い数値を示した。溶接後の熱収縮による歪みパターン予測では、標準データベースの最大240%の誤差が、独自データベースでは最大121%に改善された。開発したAIモデルによる予測結果を適用した結果、標準データベースより優れた精度で実機変形を予測できた。筐体フランジ面のZ方向の変位差で実機計測が2.11mmであるのに対して、AI予測値を活用した解析結果は1.88mmになり、高い整合性を示した(図10、図11)。

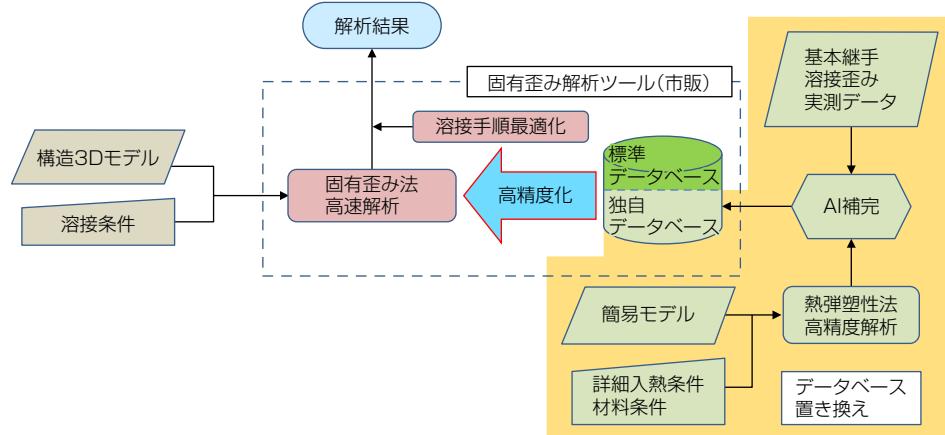


図8-データベース置き換えシステム

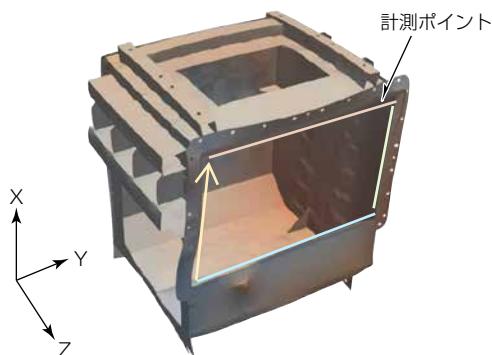


図9-溶接後の遮断器筐体解析結果

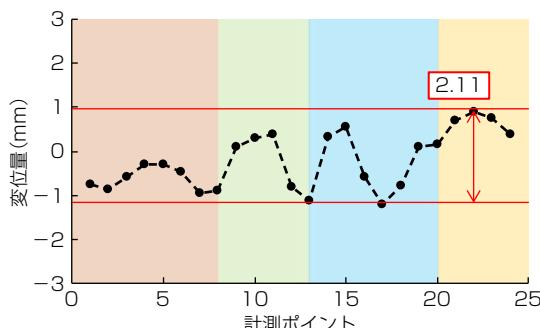


図10-実機フランジ取り合い面のZ方向変位差

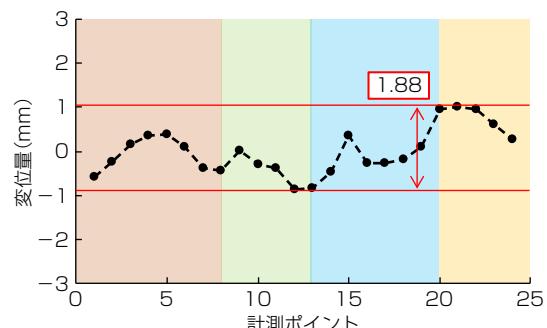


図11-AI解析したフランジ取り合い面のZ方向変位差

このケーススタディーによって、AIモデルと高精度データベースの組合せが溶接構造物の変形予測に有効であることが確認された。これによって複雑な溶接構造物の高精度変形予測が可能になり、設計段階での最適化とリードタイム短縮を実現した。

4.4 今後の展望

この研究で開発した溶接パラメーター自動最適化システムは高い実用性を示すが、更なる発展を目指している。重要課題はAIモデルの適用範囲拡大であり、今後は異なる材料や特殊環境への適用も視野に入れている。これによって熟練技術者の知見のデジタル化と技術継承を促進し、製造業の持続基盤構築に貢献する。この技術は製造プロセスの効率化と高品质製品の安定供給を実現する重要な技術として発展が期待される。

5. む す び

樹脂成形、機械加工、溶接の製造基幹工程でのAI活用技術を述べた。これらは熟練者の暗黙知をデジタル化し、製造パラメーターを自動最適化することで、品質向上と生産効率化を両立させる。

今後は、技術の精度向上と適用範囲拡大によって、製造業のデジタル変革を促進し、持続可能な生産体制構築に貢献する。



生成AIを活用した社内技術文書検索システムの構築と展開

Building and Deploying Internal Technical Document Search System Using Generative AI

石橋祐太郎*
Yutaro Ishibashi
宮下修治*
Shuji Miyashita
深井哲人†
Yoshito Fukai

*生産技術センター
†設計技術開発センター

要 旨

三菱電機は、ものづくり領域での業務効率化を目的に、ECM(Engineering Chain Management), SCM(Supply Chain Management), DCM(Demand Chain Management)の各業務領域で、生成AIを活用した社内技術文書検索システムの構築とその全社展開に取り組んでいる。具体的には、RAG(Retrieval-Augmented Generation)を基盤とした検索プラットフォームをはじめとして、ソフトウェア開発での類似障害検索システム、LSI開発を支援するRAGベースの検索システム、AIを活用したコンタクトセンター基盤などの様々なソリューションを開発し、展開を推進している。これらの取組みによって、情報探索時間の削減、手戻りの抑制などの業務効率化を実現している。

1. まえがき

急激に変化する事業環境の中で持続的成長を確保するためには、ものづくり領域でのリードタイム短縮が急務になっている。設計・製造領域では、社内外に蓄積された豊富なデータやノウハウを有効に活用することで、業務プロセスの大規模な効率化が期待できる。一方、近年、生成AIの技術が飛躍的に発展し、その活用が課題解決の大きな鍵として注目を集めている。この背景を踏まえて当社では、生成AIを活用した社内技術文書検索システムを構築し、各種技術文書やマニュアル、客先対応などのナレッジデータを横断的に検索可能にする仕組みを整備している。

本稿では、それらの取組みとその展開の状況について述べる。

2. 社内知識を探索する生成AIプラットフォーム

社内特有の技術知識を大規模言語モデルに反映させるには、ファインチューニングにかかる膨大なコストが課題になる。そこで、生成AIに検索機能を組み合わせるRAGを採用し、必要な情報を取り出して回答生成に反映させるツール“Metriever”を開発して、全社への展開を進めている(図1)。Metrieverは、前処理条件やプロンプトの柔軟な設定、図

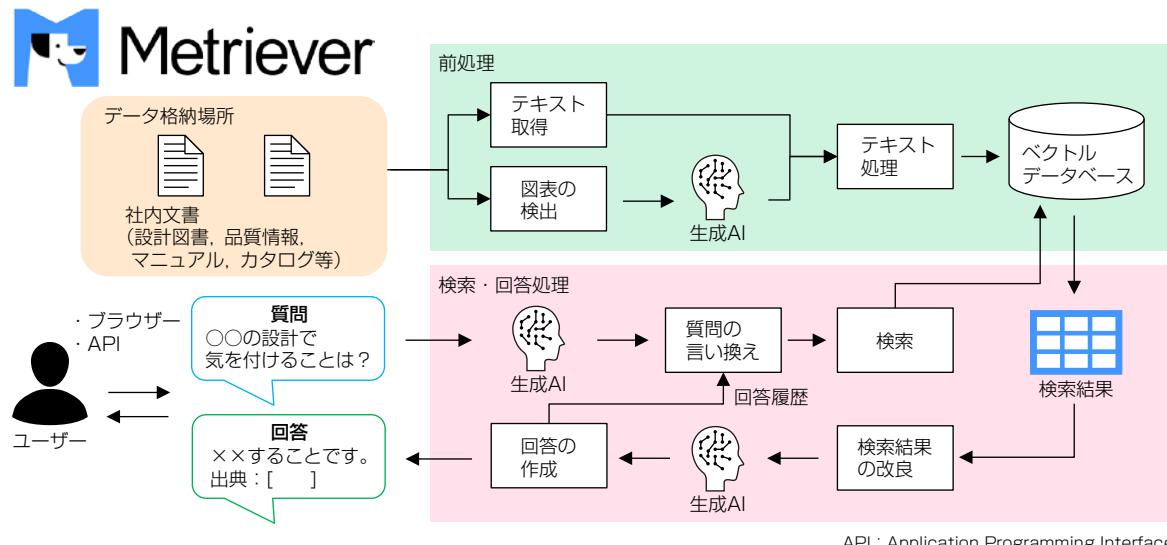


図1-Metrieverの構成と処理フロー

表の取扱いなど、多様なユースケースに応じて、細かなカスタマイズが可能である。これによって、様々な分野の開発プロジェクトなどで、利用単位での適切な知識獲得と回答を実現している。

3. ソフトウェア開発向け類似不具合検索

IoT(Internet of Things)化やクラウド活用の進行によってビジネスモデルが変化したことで、ソフトウェア開発では、新たなサービスやソリューションを高頻度で更新することが必要になっている。リリースのリードタイムを短縮するのに重要な不具合対応では、過去に蓄積された情報を効率的に参照することが求められる一方で、時間が経(た)つにつれて情報が肥大化・複雑化・老朽化し、ブラックボックス化しやすくなるという課題がある。また、そのような問題を解決する手段として、自然言語処理とAIの活用が有望視されている。過去の不具合情報を学習することで、新たに発生した原因不明の故障に対して、原因の可能性が高い障害、エラー、バグを自動的に推論し、調査範囲の絞り込みによる効率化が可能になる。

そこで、Elasticsearch^(注1)やn-gramなどの自然言語処理を用いて、単語又は自然文を基に類似する過去の障害情報を検索するシステムを開発した。ユーザー要望に対応して類義語辞書を加味した検索機能、複数カラム・複数の検索対象の切替え機能、n-gramと単語の最適な重みづけの設定、ハイライト表示機能、完全一致検索機能といった高精度な検索機能も実装した(図2)。また、この検索機能を不具合情報の管理データベースとして利用されているRedmineのプラグインにすることで、閲覧権限の制御や、追加のハードウェアリソースを抑えた低コストでの導入を実現できた。

#	ID	システム名	発生日	確... 内容	評...
1	355	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:39:59	1.00 Alert SSL設定の場合	
2	244	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:39:49	0.95 HUBのSSL設定について	
3	415	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:40:04	0.93 HUBSSL証明書設定について	
4	114	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:39:38	0.83 OSSライセンス情報サイトの運用状況について	
5	356	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:39:59	0.83 V2019.10.0でのSSL通信設定	
6	67	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:39:34	0.80 OSS名 [REDACTED]について	
7	172	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:39:43	0.80 HubによるOSS挿入不可能なパターンについて	
8	245	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:39:49	0.78 (質問) プロジェクト内OSSの脆弱性情報の取得方法	
9	480	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:40:10	0.78 結果EXCELで、ソース/OSSコードの比較表示	
10	492	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:40:11	0.75 許可済OSS情報取得のSDKについて	
11	473	三菱電機様 - [REDACTED]課題管理	2024-08-19 06:40:09	0.71 (質問) コードマッチがなくOSS判定されるケースについて	

図2-Redmineプラグイン版の類似障害検索のUI(User Interface)画面

さらに、次のステップとして、生成AIを用いたRAGを用いたシステムを構築した。これによって、ユーザーが記述したプロンプトを基に、自然言語処理で検索した結果をユーザーが解釈しやすい形に整形処理できる。このプロセスは、目的的結果を得るまで繰り返し行うことができる(図3)。このアプローチによって、類似した故障・障害情報を探すだけでなく、更に詳細な質問を行うことで、原因になるエラーやそれに関連するバグ、及びその対処方法など、ユーザーが求める形の回答を得ることが可能になった。今後、不具合情報データベースだけでなく、仕様検討時の議事録、仕様書、試験仕様書などのドキュメントを検索対象にすれば、このツールの活用範囲を更に拡張できると考えられる。

(注1) Elasticsearchは、Elasticsearch B.V.の登録商標である。

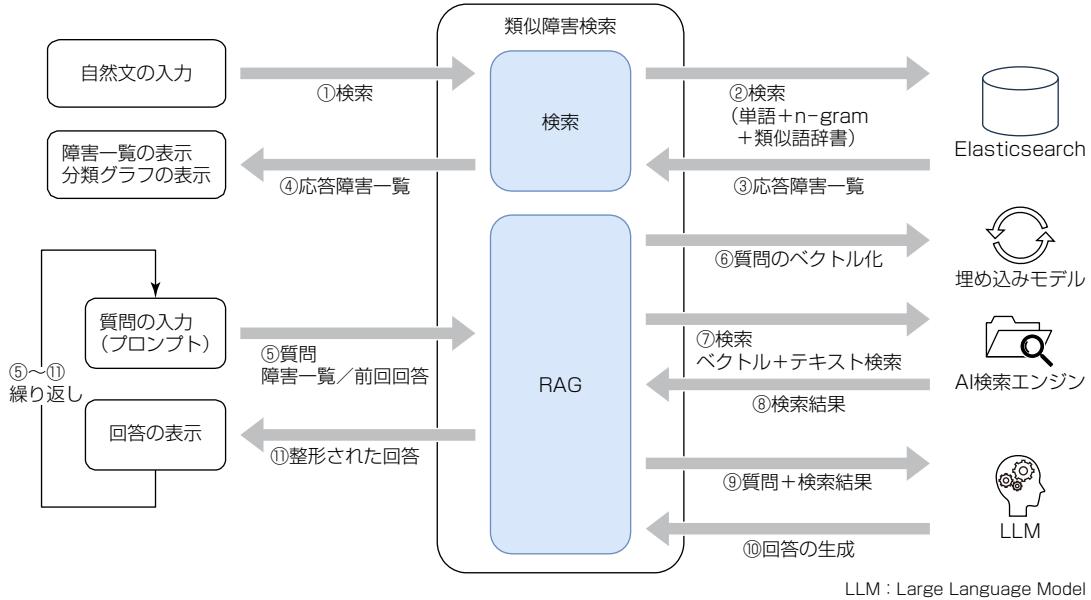


図3-類似障害検索の生成AI活用のメカニズム

4. LSI開発向けRAG

LSI開発は、製品競争の激化に伴って短納期化が求められている。LSI開発分野では、HDL(Hardware Description Language)を用いてLSIの振る舞いを定義し、ハードウェア化する。HDLは、一般的なソフトウェア言語と比べて抽象度が低いため、生産性が低く、効率的な開発手法やツールの活用を模索する必要があった。

近年、生成AIによるコーディングがあらゆるソフトウェア言語で注目されており、生成AIでHDLコードを生成できれば、生産性を大幅に高めることが期待できる。LSI開発では、生成AIの学習データにはない当社の設計基準に沿ったHDLコードを生成する必要があり、生成AIを活用する上での課題になっていた。そこで、RAGを用いて当社の設計資産である過去のHDLコードを検索して生成AIに与えることで、基準を満たしたHDLコードを生成することを可能にした(図4)。

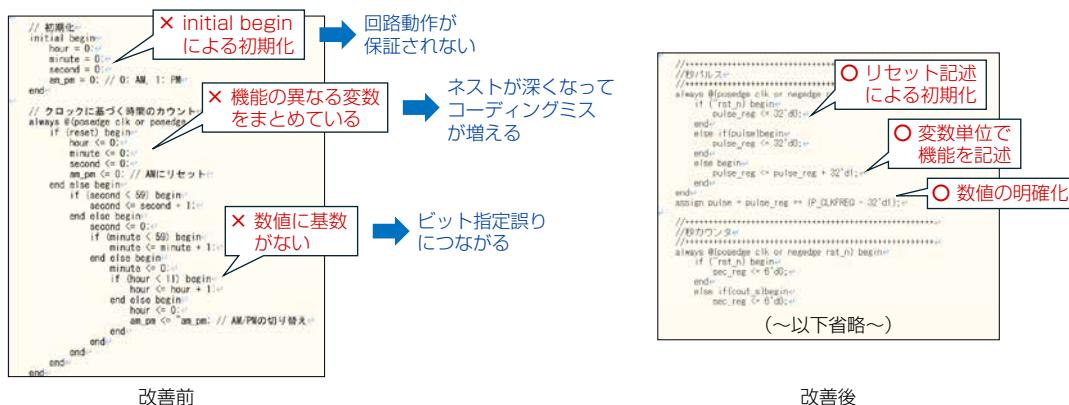
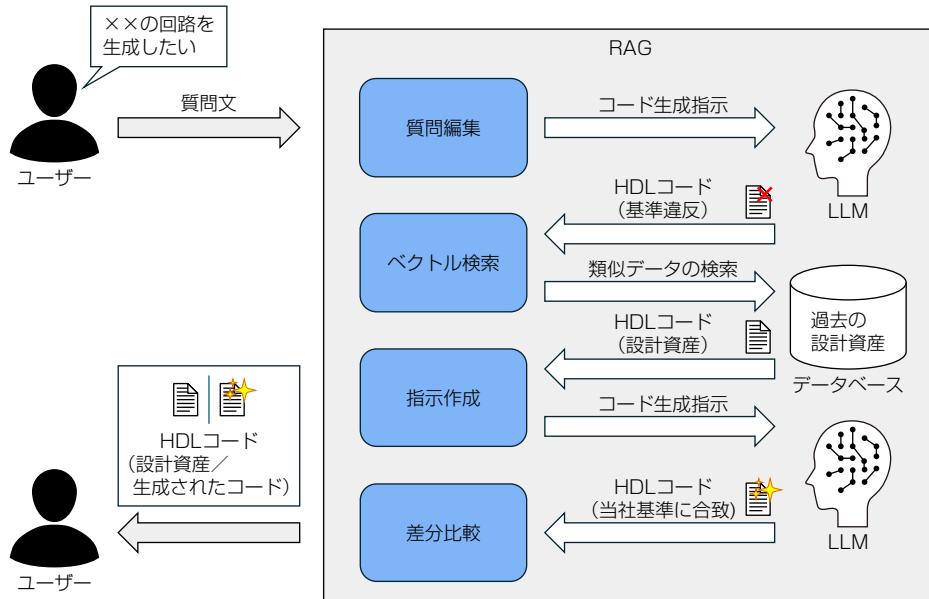


図4-RAGによるHDLコード改善例

具体的な手順について述べる。ベクトル検索でデータベースから所望の情報を抽出し、抽出した情報をLLMに参考情報として与える。ベクトル検索は、データベースに類似性の高い情報がなければ、適切な設計資産を抽出できない。そこでLLMに対して、所望の機能を持つ“HDLコード(基準違反)”を生成させて、ベクトル検索にかける。ベクトル検索の結果、所望の機能と類似性の高い、設計基準を満たした“HDLコード(設計資産)”が抽出される。改めてLLMに“HDLコード(設計資産)”を参考にして、HDLコードを生成するよう指示する。生成したHDLコードは、参考にしたHDLコードと比較し、差分を明確にしてユーザーに提供する(図5)。

なお、LSI設計へのAI活用に関するほかの取組みは、本号の特集論文“AI活用によるLSI開発の高度化”に述べている。



5. 生成AIを活用したコンタクトセンター基盤

コンタクトセンターは顧客とのコミュニケーションの最前線であり、迅速かつ正確な顧客対応が求められる。生成AIはその精度とレスポンスの速さからコンタクトセンターの業務と親和性が高いため、大幅な業務変革が期待できる。そこで、コンタクトセンターのコアになる四つの業務領域に生成AIを適用したコンタクトセンター基盤の構築に向けて、要素技術の開発、及び全社展開を推進している(図6)。

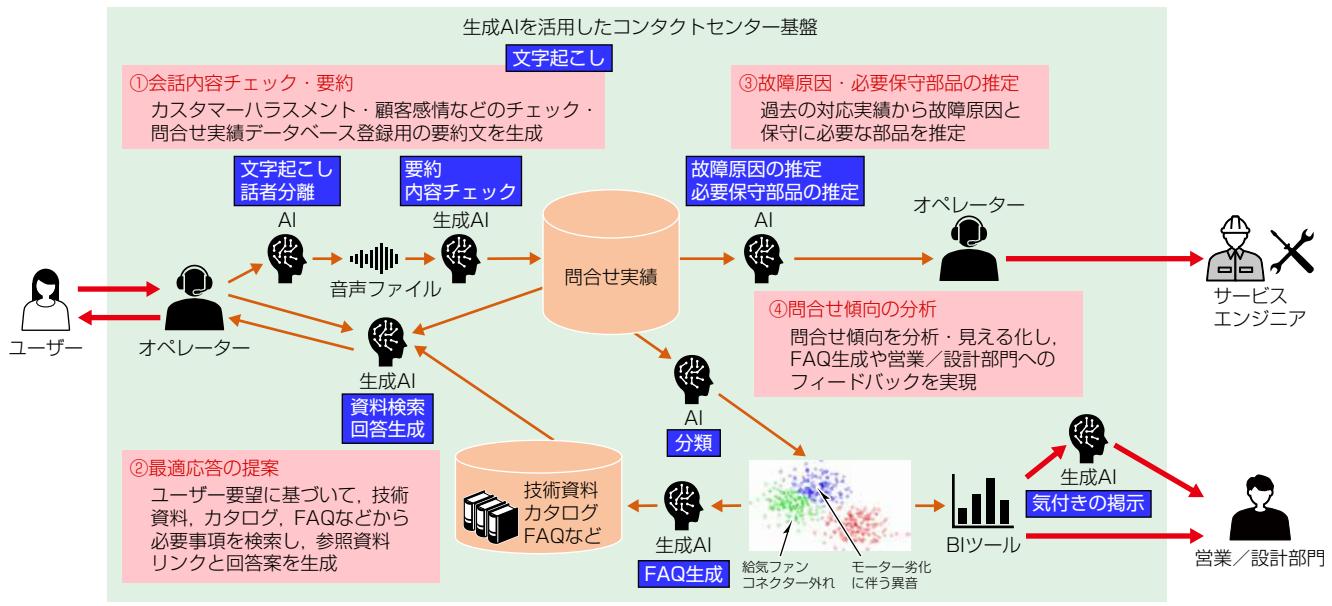


図6-生成AIを活用したコンタクトセンター基盤

この基盤では、最適応答の提案と問合せ傾向の分析に生成AIを用いた検索システムを活用している。最適応答の提案では、2章で述べたMetrieverを用いて技術資料、カタログ、FAQから顧客の質問に対する回答を生成している。ただし、販売店や施工店からの問合せは質問内容が非常に複雑であり、複数の資料を参照して数値計算などを行った上で回答

する必要がある。例えば、精度が高い回答が得られるマニュアルをベースとした回答業務では回答案まで作成し、複雑な演算が伴う業務では必要情報検索の用途として用いるなど、各業務に合った使用方法を定めている。また、現状回答案作成が難しい複雑な用途については、AIエージェントの活用で実現することを検討している。

問合せ傾向の分析では、問合せ実績を分析してFAQの候補を抽出する。コンタクトセンターに寄せられる様々な問合せの全てを人手で分析することは現実的ではないため、クラスタリングした問合せ実績と現在のFAQとの類似度を自然言語処理AIで算出し、問合せは多いがどのFAQにも類似していない問合せをFAQ候補としてピックアップする。さらに、生成AIを用いてこれらのFAQ候補となる問合せ実績から、新規FAQの生成や既存FAQの更新を行う。

6. む す び

生成AIを用いた検索システムを用いて社内知識を有効活用し、業務を効率化する取組みについて述べた。今後は開発したツールの全社展開を通じて抽出した現場ニーズを基に、最新技術の適用によってニーズを満たす機能開発と更なる精度向上を図っていく。さらに、ものづくり領域での生成AI活用の対象業務を拡大することで、業務の効率化と開発リードタイム短縮に貢献していく。

参 考 文 献

- (1) 玉谷基亮：三菱電機グループの持続的なものづくりを支えるAIソリューション群、三菱電機技報、98、No.9、3-01～3-05 (2024)



特集論文

DevOps×生成AI×セキュリティーを搭載した ソフトウェア開発基盤“MelSecOps”

Development Infrastructure Equipped with DevOps, Generative AI and Security “MelSecOps”

尾崎紀之*

Motoyuki Ozaki

長峯 基†

Motoi Nagamine

小森裕之†

Hiroyuki Komori

梅原友幸*

Tomoyuki Umebara

*設計技術開発センター

†電力システム製作所

‡設計技術開発センター(博士(情報科学))

要 旨

近年、事業環境が大きく変化し、需要予測に基づく見込み生産から市場環境の変化に追従できるフレキシブルなものづくりへの変革が求められており、開発リードタイムの短縮が重要な事業課題の一つになっている。しかし、多くのソフトウェア開発現場では技術的負債の蓄積や開発環境の陳腐化が足かせになり、現場独力では改善が進まない現状があった。そこで三菱電機は、DevOps^(注1)、生成AI、セキュリティーなどに関する各種ツールを備えた当社グループでの統一したソフトウェア開発基盤“MelSecOps”を構築しグループ全体に展開することで、開発リードタイムの短縮を推進している。

(注1) ソフトウェアの開発リードタイム短縮・品質向上等のために、開発担当者と運用担当者が協力する手法や仕組み

1. まえがき

近年、顧客ニーズの多様化、技術革新の加速、政策・規制の変動など、市場環境の変化が激しくなっている。このような状況の中、当社グループでは、グループ内外の知見の融合と共に創によって、進化した統合ソリューションを提供する循環型デジタル・エンジニアリング企業へ変革し、多様化する社会課題の解決に貢献することを目指している。そのためには様々なデータを集め・分析し潜在的な顧客要求を把握することと、要求を製品やサービスなどの形で早期に具現化することが重要である。このように、新製品・新サービスをいち早く市場に投入し、市場からのフィードバックを受けて迅速に適応することは顧客価値の最大化につながる。

これらの理由から、開発のリードタイムを短くすることは、重要な事業課題の一つになっている。しかし、当社グループ内の多くのソフトウェア開発現場では改善が進んでいない現状がある。これは、ソフトウェア開発業務の負荷が高く、自動化等の改善や新技術の導入に向けた時間確保が困難なためである。また、一度自動化等の改善の仕組みを導入してもそれを保守・維持する負担が重く、開発環境が陳腐化してしまっている現場もある。これらの課題を解決するため、当社グループでの統一したソフトウェア開発基盤MelSecOpsをコーポレート組織で構築し当社グループ全体に提供することで、ソフトウェア開発のリードタイム短縮を図る。

本稿では、MelSecOpsで提供する各種サービスの概要及びソフトウェア開発現場への適用事例を述べる。

2. MelSecOpsの概要

この章では、MelSecOpsが提供する各種サービスを述べる。MelSecOpsは当社グループでの統一したソフトウェア開発基盤であり、DevOps、生成AI、セキュリティーの三つのサービスで構成される。これらのサービスには、当社グループ独自のノウハウや工夫を織り込んでいる。

2. 1節では、サービスの全体像を述べる。2. 2節～2. 4節では、それぞれDevOps、生成AI、セキュリティーについての取組みを述べる。

2.1 全体像

図1にMelSecOpsが目指すソフトウェア開発の姿を示す。MelSecOpsでは、ソフトウェア開発の各種作業を支援するツールを提供することで、開発リードタイムの短縮を狙っている。支援対象は設計・実装・テスト等のエンジニアリングプロセスに限らず、プロジェクト管理を支援するツールも提供する。例えば、進捗・課題・品質の可視化や生成AIによる支援を通じて、プロジェクト状況のタイムリーな把握や迅速な意思決定をサポートする。

これらの支援を実現するためのMelSecOpsのシステム構成を図2に示す。MelSecOpsはクラウドベースの開発基盤で

あり、リポジトリ、DevOps(チケット管理、ダッシュボード、パイプライン)、生成AIやセキュリティーツールから成る。さらに、オンプレミス環境にあるツールとの連携やナレッジベースのような他環境との連携も担う。クラウドベースにした主な理由は、サーバー等の準備が不要なためユーザー(グループ内のソフトウェア開発部門)に迅速に環境を提供できること、従量課金のため初期投資を小さく始められることの二点である。各ツールは個々に独立して利用可能であり、ツールの価値を更に高めるために、ツール間の連携機能も開発予定である。具体例は各項目で述べる。

2. 2節以降で、各サービスとそれを実現するツールの詳細を述べる。

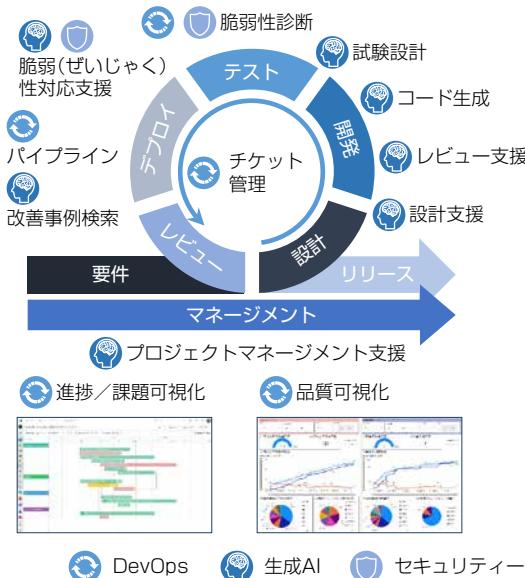


図1-MeSecOpsが目指すソフトウェア開発の姿

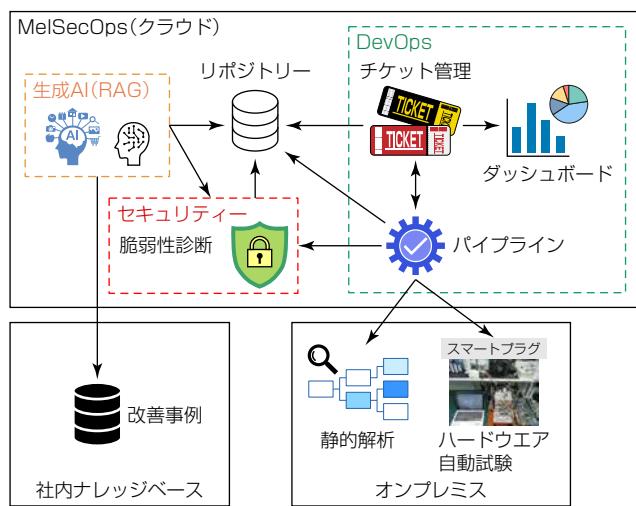


図2-MeSecOpsのシステム構成(概念レベル)

2. 2 DevOps

DevOpsでは次のツールを含むSaaS(Software as a Service)^(注2)を提供する。また、開発部門がSaaSを利用するに当たって、初期設定等の環境導入のハードルが高いという事情から、標準・推奨設定やサンプルを併せて整備した。

(注2) サービス提供事業者のサーバーでソフトウェアが動作し、そこにユーザーがアクセスして利用する形態のアプリケーション

2. 2. 1 チケット管理

タスクや障害対応の進捗や工数などを記録するツールである。当社グループではタスク管理や問題管理に表計算ソフトウェアを利用している開発現場が多く、プロジェクト管理が非効率であるという問題があった。チケット管理システムを活用することで、プロジェクト全体と詳細なタスクの進捗把握の両立が容易になる、変更履歴を追跡できる、変更時に通知を受信できる、などの利点が得られる。一方、チケット管理システムを新規に導入するには、チケットの設定(入力項目やステータスの遷移など)の検討が必要であり、敷居が高いという課題があった。そこでMeSecOpsでは、システム自体に加えて、チケットの定義、権限設定、進捗把握のための表示設定(例: 遅延の警告表示)などを含む標準設定及び設定を適用するための補助ツール・手順書を提供する。これらは、当社グループのソフトウェア開発現場でのベストプラクティス等のノウハウを基に作成している。例えば、障害対応のチケットには検出工程、混入工程、混入原因、手戻り(対応)工数など、ソフトウェアの品質評価や改善施策検討のために開発現場でよく用いられている情報を入力項目に含めた。同様に、レビューのチケットにはレビュー対象の分量とレビュー時間、その子チケットであるレビュー指摘対応のチケットには指摘への対応工数の入力項目を設けた。

2. 2. 2 ダッシュボード

チケット管理システムに入力した情報を集計し、プロジェクトの進捗状況やソフトウェアの品質情報を可視化するツールである。チケット管理システムでも一部の集計や可視化は可能だが、プロジェクト管理のための十分な機能は備わって

いなかった。そこで、当社グループのソフトウェア開発でよく使用される指標・図表を自動で集計・表示する仕組みを構築した。2.2.1項のチケット管理システムで述べた標準設定を利用している場合、カスタム不要でそのまま利用できる形で提供する。例えば、レビュー指摘や障害による手戻り工数、テスト密度・障害密度、障害の検出・解消件数のデイリーでの変化、タスクの進捗を予定と実績で比較するガントチャートなどの集計が可能である。

2.2.3 パイプライン

ビルド、テスト、デプロイ等の定型的な作業を自動化するツールである。自動化が進んでいない開発現場や、自動化環境は存在するがサーバー等の保守が負担になっている開発現場を対象に、クラウドベースの自動化の仕組みを提供する。パイプラインを利用するためには専用のスクリプトを記述する必要があり、開発現場で導入する際の障壁になり得るという課題があった。そこで、パイプラインで自動実行することが多いと見込んだ処理について、スクリプトのカタログを提供する。カタログでは、再利用しやすい粒度でスクリプトのテンプレート(部品)とその利用方法をまとめた。ユーザーが自動化したい内容に応じて、カタログに掲載した部品を選択し組み合わせることで容易にパイプラインを構築できるようにして、導入の敷居を下げた。カタログに掲載した部品の例を表1に示す。

表1-パイプライン部品例

パイプライン部品名	内容
ビルド	アプリケーションをビルドする
単体試験	単体試験を実施する
コード静的解析	コードの静的解析を実施する
成果物リポジトリ登録	成果物をリポジトリ登録する
デプロイ	成果物をデプロイする

パイプラインは、開発現場が持つ既存の自動化環境と統合する機能を持つ。2.1節で述べたとおりMelSecOpsはクラウドベースでサービスを提供するが、開発現場が持つ自動化環境はオンプレミスに構築することが多い。既存の自動化環境をクラウドに移行するためには追加の開発コストを要するため、オンプレミスの環境をそのまま利用できることに対するニーズは高い。また、ハードウェアを伴う自動試験環境や、ノードロックのライセンス(特定の機器での使用が許可されるライセンス形態)で持っているツールなど、クラウドへの移行が困難な環境も存在する。このような既存の環境を有効活用するために、クラウドとオンプレミスを相互に接続し連携可能な機能を持つツールを選定した。

また、ソースコードリポジトリと連携し、ブルリクエスト(あるブランチでの変更内容をほかのブランチに反映(マージ)する要求)を発行する際にパイプラインの実行を強制する機能も持つ。これは、大人数の開発では各開発者によるソースコードの変更をマージする際に問題が発生する可能性が高く、マージによる誤りの混入を防ぐためである。この機能を活用することで、開発規模と品質確保を両立しやすくなる。

これらのほかに将来的には、パイプラインを用いてチケット管理システムに登録したチケットの情報を定期的に生成AIで分析することで、プロジェクトの健全性チェックや遅延・リスクの兆候把握に活用することを構想している。

2.3 生成AI

ソフトウェア開発の効率を劇的に改善するために、生成AIの活用ニーズは急速に高まっている。特に、長期間継続する事業が多い当社グループでは、これまでの製品・サービス開発で蓄積したナレッジを有効活用できるRAG(Retrieval-Augmented Generation：検索と生成AIを組み合わせて回答精度を向上させる技術)の需要が多くある。一方、各組織に生成AIを用いたアプリケーションを開発できるエンジニアが必ずしもいるわけではないため、組織ごとに開発することは非効率である。そこで、MelSecOpsでは生成AIを活用する標準化したアプリケーションを開発し(表2)、プロンプトの調整やRAGで参照するデータベースの作成だけユーザーごとにカスタマイズすることで、即座に生成AIを用いたアプリケーションを開発組織に導入できる仕組みを構築した。これによって、生成AIを活用したアプリケーションの開発スピードを大幅に向上させることに成功した。

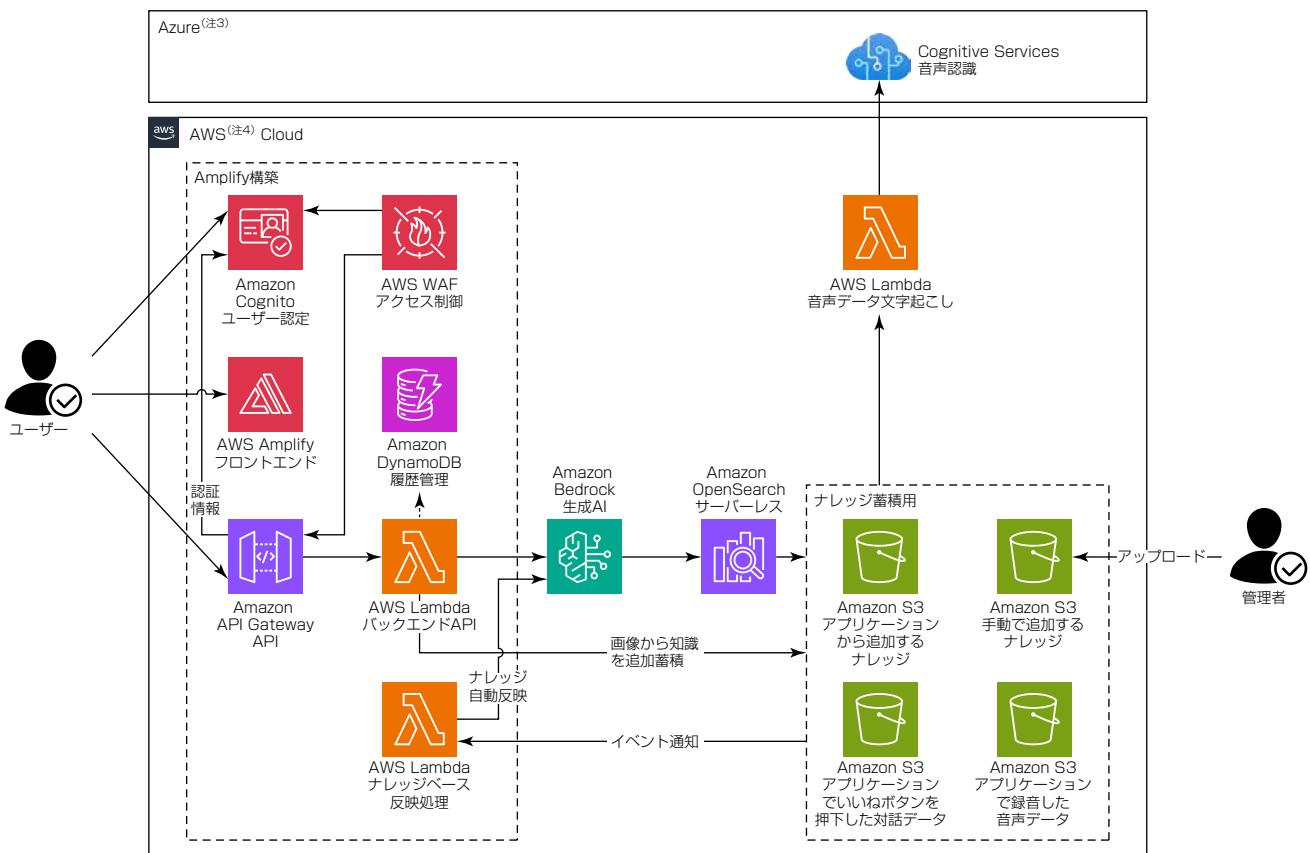
生成AIを活用したアプリケーションの中で、仕様問合せツールのアーキテクチャーを示す(図3)。このツールは、RAG技術を使用しており、RAG技術を用いるアプリケーションの基本構成は同様のアーキテクチャーを採用している。RAG技術を活用したアプリケーション(以下“RAGアプリ”という。)の特徴として、精度向上のために二つの機能を設け

ている。一つ目はナレッジアップロードである。この機能は、ユーザーがアプリメニューから個別にナレッジを追加できるようにすることで、不足しているナレッジをユーザーが追加し、精度を向上させることを可能にするものである(図4)。二つ目は“いいね機能”である。この機能は、RAGアプリへの問合せ結果に対して“いいね”ボタンで応答できるもので、このユーザーの応答情報を次回の問合せ時に活用することで、有効な回答を優先的に生成することを可能にしている。

また、改善事例を格納したグループ内ナレッジベースを生成AIで椻索するツールも持つ。当社グループでは、数多くのソフトウェア開発現場の知見を有効活用するために、開発でのベストプラクティスをグループ内に広く共有するWebサイトを持っている。しかし、サイトの検索性が問題で利用者が伸び悩んでいた。そこで、探したい情報をプロンプトで入力すると当該Webサイト内を椻索し、蓄積したナレッジの中から類似している情報を提示するRAGアプリを構築した。RAGアプリの公開を含めたリニューアルの結果、サイトの利用者数・訪問回数共に約2倍に増加した。

表2-MeSecOpsが提供する生成AIを活用したツールの一覧(開発中、開発予定のものを含む)

ツール名	内容	効果
作業見積り	過去案件の工数実績等を参考に、作業工数を見積もる	・プロジェクトの遅延リスクの低減 ・再計画のリスク低減
仕様問合せ	仕様書やマニュアルに記載されている文章の要約。問合せへ自動応答する	・ドキュメント確認の工数削減 ・問合せ対応の工数削減
設計書検索	開発内容から変更が必要になる対象機能を記載した設計書をデータベースから検索する	・設計検討の工数削減 ・ナレッジ共有の効率化
レビュー支援	過去のレビュー指摘内容を基に、ドキュメント(設計書、仕様書)をレビューする	・レビューの工数削減 ・不具合の再発防止
試験設計	仕様書を基に、試験ケース・試験手順を生成する	・試験設計の工数削減 ・テストカバレッジ(網羅率)の向上
脆弱性対応支援	ソースコードを解析し、解析の結果から脆弱性に対するソースコードの修正案を提案する	・脆弱性対応の工数削減
プロジェクトマネージメント支援	過去のプロジェクト実績・事例と現在のプロジェクトの状況を基に、リスクを抽出する	・リスクの早期発見 ・遅延の防止
改善事例検索	当社グループのWebサイトを対象に、ユーザーにマッチするソフトウェア開発の改善事例を検索する	・良好事例の展開促進



(注3) Azureは、Microsoft Corp.の登録商標である。
(注4) AWSは、Amazon Technologies, Inc.の登録商標である。

図3-仕様問合セツールのアーキテクチャー



図4-仕様問合せツールの画面例

2.4 セキュリティ

近年、サイバー攻撃の多様化、ネットワークに接続する組み込み機器の増加、サイバーセキュリティに関する各国の法令・規制の整備などに伴って、製品のセキュリティ確保のニーズが高まっている。そこでMelSecOpsでは、脆弱性診断ツールを先に述べたパイプラインで自動実行する機能を提供する。当該ツールでは、ソースコード、コンテナの設定、依存関係のあるオープンソース等に脆弱性が含まれていないかを自動で確認できる。パイプラインを活用しソフトウェアのビルドやテストと一緒に脆弱性を診断することで問題を早期に発見できるため、セキュアな製品・サービスの提供につなげることができる。また、チケット管理システム及びダッシュボードと連携し、脆弱性への対応状況を管理することも可能である。

さらに、脆弱性診断ツールによる指摘内容を生成AIに入力し、ソースコードの修正案を提示することで脆弱性対応を効率化できないか技術検証中である。脆弱性診断ツールが出力する解析結果であるSARIF(Static Analysis Results Interchange Format)ファイルに含まれる膨大な情報は、人間が理解して修正対応をするのに工数がかかるという問題がある。そこで、ソースコードとSARIFファイルを特定のプロンプトとともに生成AIに入力することで、指摘内容に補助的な情報を付与して理解を助けたり、ソースコードの修正案を提示したりできないかを検討している(図5)。現在、プロトタイプによる検証が完了し、特定の事例に対して情報の付与や修正案の提示が可能であることが分かった。今後は、実製品のソースコードを対象に実証した上で、グループ内で広く活用できる汎用的なツール化を目指す。

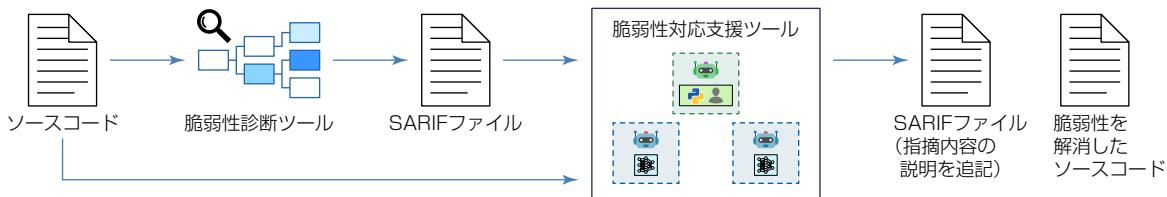


図5-生成AIを活用した脆弱性対応支援ツール

3. 導入支援サービス

この章では、当社グループの開発現場で、開発効率が低下することなくスムーズにMelSecOpsを導入できるサービスについて述べる。

3.1 サービス概要

当社グループのソフトウェア開発では、各製作所で独自の開発プロセスやルールが存在する。このため、MelSecOpsを現場主導で導入すると、現場の負担が大きくなり、導入に時間を要することが予想された。そこで、導入を効率的に進めるために、開発現場のプロセスや課題を個別に分析し、効果を迅速かつ最大限に享受できる形で導入する専属チームを

作って、専属チームによる導入支援サービスを提供した。

また、MelSecOpsは、DevOps、生成AI、セキュリティーの主要ツールを全て導入する必要はなく、個別にDevOpsだけや生成AIだけなど、取捨選択して導入することも可能にしている。この導入ポリシーの意図は、MelSecOpsを全て導入しようとすると、開発現場で混乱が生じて、導入までの準備や検討に多くの時間を要することが考えられるためである。このようにツールの部分導入を可能にすることで、利用価値をスピーディーに判断できるようにして、段階的に導入範囲や導入ツールを拡大することを狙いとしている。

3.2節以降で、二つの製作所にそれぞれ必要なツールを導入した際のサービス事例について述べる。

3.2 DevOps導入事例

この節では、MelSecOpsの主要サービスであるDevOpsを導入した事例について述べる。この導入事例の開発現場では、元々ウォーターフォール型の開発プロセスを採用しており、顧客の開発要求に迅速に対応できず、市場への製品リリースに時間がかかるという課題を抱えていた。この課題を解決するために、DevOpsを導入したものである。

DevOpsの導入に際して、現行の開発プロセスを詳細に分析し、DevOpsの効果を最大限に享受するために、開発全体はウォーターフォール開発を維持しつつ、各開発項目に対する設計から実装及び単体試験はアジャイル開発を採用するハイブリッドな開発プロセスを設計した。このハイブリッドな開発プロセスを実現するために、各開発フェーズでのコードの構成管理と実行されるパイプラインを定義した“プランチ戦略”を設計した(図6)。このプランチ戦略によって、個別の開発要求ごとの実装作業を並列して進めることができになり、ソフトウェア全体を組み合わせた総合試験の評価フェーズでも、機能ごとに品質を担保したソフトウェアを集約することで効率的な開発が可能になった。

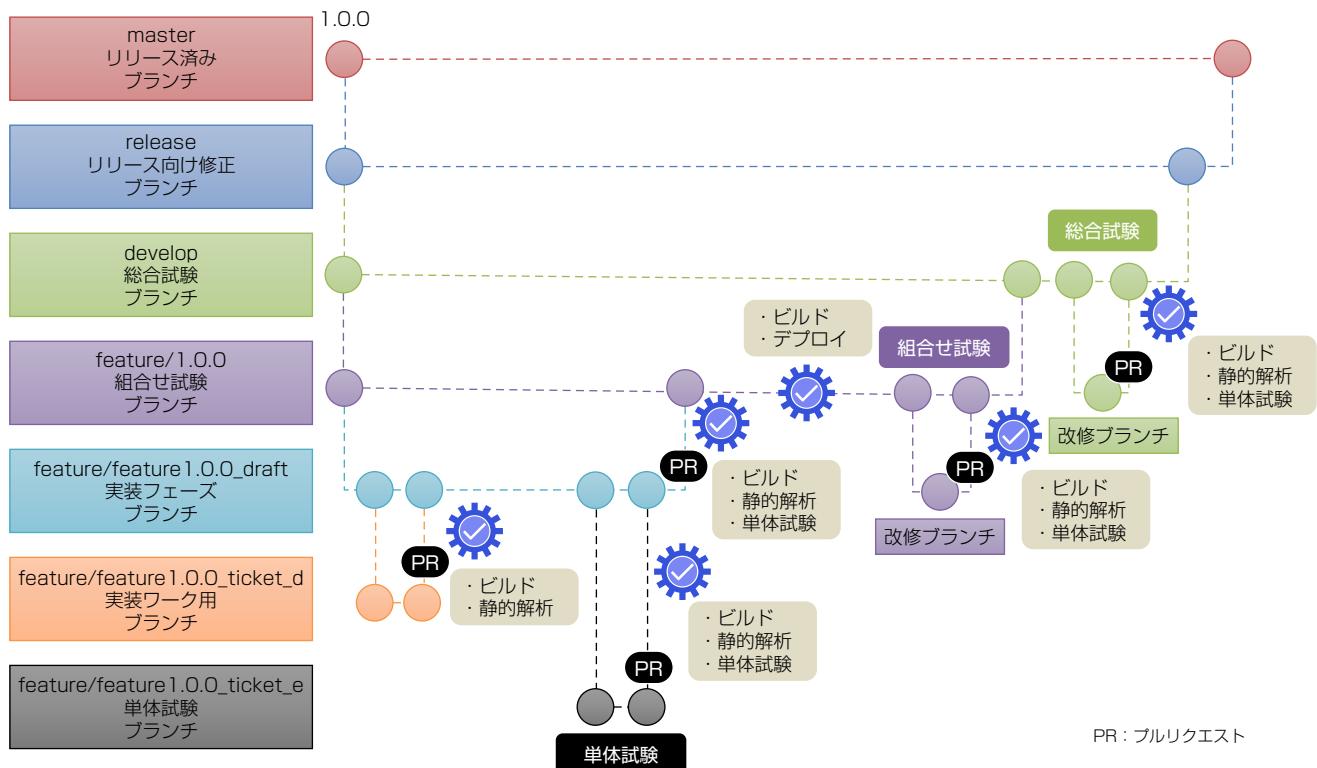


図6-開発プロセスに応じたプランチ戦略

このプランチ戦略を支援するため、2.2節で述べたパイプライン部品を組み合わせることで当該製作所向けのパイプラインを整備した。例えば、コードの修正をトリガーにビルド、コードの静的解析、単体試験を実行するパイプラインを準備している。

これによって、開発チームはDevOps上に開発環境を移行し、開発要求事項をチケットで管理し、ソフトウェア開発でのコードのビルド、静的解析、単体試験をパイプラインで自動実行することで、開発工程を効率化した。

3.3 生成AI導入事例

この節では、MelSecOpsの主要サービスである生成AIを導入した事例について述べる。この導入事例は、組み込み機器開発で、制御仕様を変更する際に、複数の制御仕様書から変更内容及び影響を受ける制御仕様を抽出するのに時間がかかっていた問題を解決するものである。この問題を解決するために、表2に示す生成AIを活用した仕様問合せツールで制御仕様書を横断的に検索するRAGアプリを導入した。

一般的に、RAGアプリの導入では、導入後に現場で活用可能な精度が出ないことが課題になるため、導入支援サービスでは業務で活用可能なレベルまで精度を向上させている。具体的には、RAGアプリが参照するデータベースに格納されるファイルなどの情報をそのまま格納するのではなく、生成AIが目的に応じて扱いやすい構造化データに前処理を行ってから格納する。構造化データにする前処理の一例として、RAGアプリが参照するPDF(Portable Document Format)ファイルはOCR(Optical Character Recognition)^(注5)で処理した上でマークダウン形式に変換し、さらに参照するデータの目的に応じてチャネルサイズを調整することで、生成AIが回答を導出しやすくなる(図7)。これによって、RAGアプリの正答率が50%以上改善した。さらに、ユーザーインタビューによる定性評価と、RAGAS(RAG Assessment System)による定量評価によって、プロンプトチューニングを繰り返すことで、検索の正答率を当初の50%から70%以上に向上させた。

(注5) 画像から文字を認識しテキストデータに変換する技術

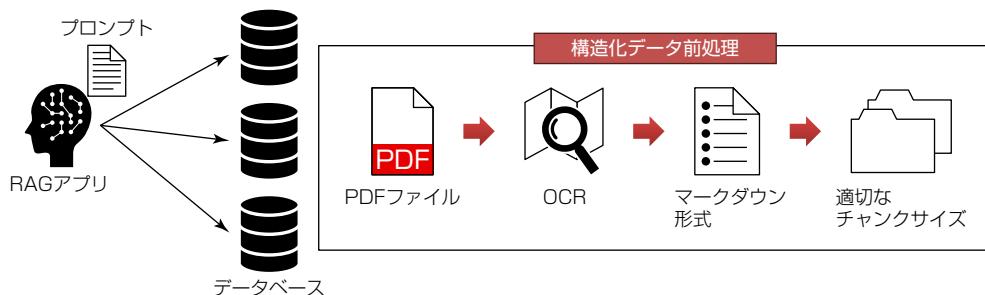


図7-構造化データ前処理例

4. むすび

当社グループでの統一したソフトウェア開発基盤であるMelSecOpsについて述べた。MelSecOpsではソフトウェア開発の各種作業を支援するツール群を提供することで、ソフトウェア開発のリードタイムを短縮できる。

今後は、提供するツールの拡充及び既存ツールの機能追加を通じて、更なるリードタイムの短縮を狙う。また、MelSecOpsが開発現場に普及するためには、効果的なツールを提供するだけではなく、導入をサポートすることも重要である。そのため、3章で述べたような、開発現場と伴走し、課題把握、改善の提案、開発の特性に合わせたMelSecOpsの設定・カスタマイズ等を行う導入サービスを拡充する。

量子暗号

Quantum Cryptography

*情報技術総合研究所(博士(理学))

要 旨

量子暗号(Quantum Cryptography)は、量子技術を利用した暗号方式であり、絶対に破れない暗号方式として知られる。三菱電機は2000年前後から量子暗号の研究を開始し、装置実装と理論の両面で様々な成果を上げており、現在は理論研究に注力している。理論研究というと、実験結果を後追いで述べるだけと思われがちだが、量子暗号の場合は必ずしもそれにとどまらず、現実の量子暗号装置の性能向上やコスト削減に直接役に立つものである。

1. まえがき

本稿ではまず、量子暗号とは何かを、既存の暗号方式と比較したのち、この分野での当社の研究成果を述べる。日常生活で最も馴染(なじ)みのある暗号は、無線LANでのパスワード暗号化だと思われるため、それと比較する。三菱電機は2000年前後から量子暗号の研究を開始し、装置実装と理論の両面で様々な成果を上げてきたが、現在は理論研究だけに注力している。そのため、理論研究を中心に成果を述べる。

なお本稿での量子暗号とは、専門的には量子鍵配達(Quantum Key Distribution : QKD)のことだが、一般には量子暗号という呼び方が定着しているので、ここでもそれを採用した。また、詳細を知りたい場合は参考文献(1)を参照されたい。

1.1 既存の暗号方式(現代暗号)との比較

量子暗号の目的は、無線LAN暗号化をはじめとする既存の暗号方式(以下“現代暗号”という。)と全く同じで、ユーザーが選んだメッセージを、誰にも盗聴されずに(以下“安全に”という。)誰かに送ることである。

1.1.1 暗号通信での役割が異なる

無線LAN暗号化の場合は、ユーザーが一旦手動でパスワードを設定すれば、それ以降はルーターが通信内容を暗号化し続ける。暗号の使い方は基本的に全てこれと同じであり、抽象的には、次の2ステップに分かれる。

ステップ1 “秘密鍵配付”：秘密の基になる、ランダムな文字列を何らかの方法で配付する。その文字列が覚えられるほど短ければパスワード、長くて完全にランダムであれば秘密鍵と呼んで区別することが多い(以下“秘密鍵”という。)。

ステップ2 “暗号アルゴリズム”：ユーザーが自由に選んだメッセージを、ステップ1で配付した秘密鍵とともに、“暗号アルゴリズム”というソフトウェアに入力して暗号文を作り、相手に送る。相手も同じ秘密鍵を知っている(かつ復号アルゴリズムは公開されている)ので、メッセージを復号できる。結果として通信が成立する。

この二つのステップの中で、現代暗号と量子暗号は、それぞれ異なる役割を果たす。

まず現代暗号とは、ステップ2の暗号アルゴリズムのことである。それを実際に使うためには、あらかじめステップ1で秘密鍵を配付しておく必要がある。通常は、秘密鍵を人がその場に赴いて設定したり、媒体に記録して運んだりして配付する。このやり方をここでは“秘密鍵を手で配付する”と呼ぶことにする(なお公開鍵暗号でも、この状況は緩和されるが解消されず、証明書は手で配付する必要がある)。手で配付するので、秘密鍵を大量に又は頻繁に送ることはできない。例えば無線LANの場合、短め(数百～数千ビット)の秘密鍵を一旦設定したら数年は使い回すことになる。

一方、量子暗号は、ステップ1の秘密鍵配付を、光通信によって桁違いに高速化するものである(図1左)。典型的には数Mbpsで配付できる。このため秘密鍵を頻繁に更新できるようになり、安全性(暗号の強度)が向上する。

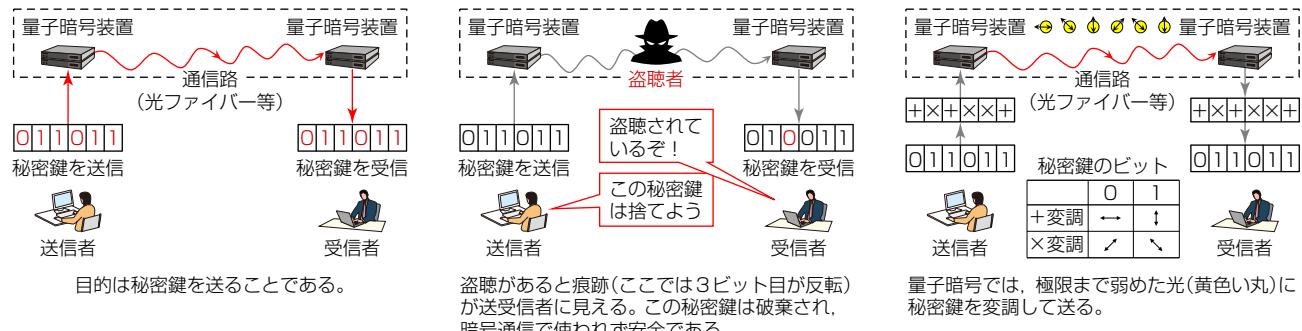


図1-量子暗号のイメージ

特に、秘密鍵の更新頻度を究極まで高めて、メッセージと同じ長さの秘密鍵を毎回使い捨てする場合、“ワンタイムパッド”という究極の暗号方式になる。

ワンタイムパッドは、従来の“手で配付する”やり方では実用的でないが（メッセージを手で運ぶのと同じことになってしまふので）、量子暗号があれば実用的になる。量子暗号は、ワンタイムパッドを実用化するための方式とも言える。

1.1.2 安全性の実現方法が根本から異なる

端的に言えば、現代暗号は“見られても漏れない”ようにするが、量子暗号は“見られたことが分かる”ようにする。

現代暗号は、ステップ2のメッセージを漏らさないようにする。そのために暗号アルゴリズムが、メッセージを複雑に攪拌（かくはん）して、見た目のかけ離れた暗号文を作る。万が一暗号文を見られても、元のメッセージを推測できないようにする。

一方、量子暗号は、ステップ1の秘密鍵を漏らさないようにする。そのために、まず“秘密鍵が通信途中で盗聴されたら、痕跡が必ず残る”仕組みを埋め込んでおく（図1中央）。そして盗聴の痕跡が見つかった場合、その秘密鍵は誰かに漏洩（ろうえい）したと結論付けて捨てて（秘密鍵はランダムな文字列であり、メッセージではないので捨てて構わない）、別の秘密鍵を送り直す。もし痕跡が見つかなければ、盗聴されていない、つまり安全だと結論付けてステップ2を使う。

量子暗号ではこの“痕跡が必ず残る”仕組みを実現するために、秘密鍵をわざと壊れやすい媒体に載せて送る。それが“量子”である。ここで量子技術が必要になる。

量子暗号の場合、量子とは、強度を極限まで弱めた光のことである（図1右）。光の強度を極限まで弱めたものを、極めて高感度で測定する場合、高校や大学で習った電磁気学からのずれが目立つようになり、代わりに量子力学の性質が見えるようになる。その性質には“光は波であるにもかかわらず、一つ二つと個数を数えられる”や“測定すると状態が変化することがある”などがある。量子暗号では、この二つ目の性質で起こる変化を盗聴の痕跡として使う。

1.1.3 実装難易度は量子暗号の方が高い

実装の難易度は、現代暗号より量子暗号の方が高い。暗号アルゴリズムはソフトウェアなので、スマートフォンやパソコン上で容易に実装できる。量子暗号は、極限に弱めた光で通信する必要があるので、専用の光通信装置が必要になる。

1.1.4 暗号強度（安全性）は、量子暗号が比較にならないほど強い

現代暗号であっても量子暗号であっても、今の暗号は強度（安全性）が非常に強くなっています。現実的な時間（年単位）で実際に解いてみせるのは難しい。したがって、各暗号方式を実際に解いてみて、各々の強度を比べるのは現実的でない。

そこで今の暗号業界では、各暗号方式の安全性を“証明すること”によって比較する。“証明すること”の具体的な作業は、安全性を数学的に証明して、それを文書として書き下すことである。良い文書が付属した暗号方式ほど強い方式とされ、ユーザーに安心して使ってもらえる。いわば暗号方式の“血統書”的に機能する。

この観点では、量子暗号が現代暗号とは比較にならないほど強い。予想に頼らず厳密な証明が書けるからである（図2）。

現代暗号は、1.1.1項で述べたとおり、短い秘密鍵を長い間使い回す前提になっている。このため最初から明確な弱点（秘密鍵を全数探索すれば必ず解ける）があり、そのままでは証明を完結できない。そこで通例は、何らかの予想（最も有名なのは“素因数分解を解くのは難しい”というもの）を持ち込んで、“もしその予想が正しかったら”という前提で証明を書

く。しかしその予想は証明できていないので、証明があるとはいえるが、結局は“現代暗号は“恐らく”破れない”としか言えない。ある日突然、新しい解読法によって破られる可能性は否定できない（なお、この状況は耐量子計算機暗号でも同じで、“耐量子計算機暗号は、量子コンピューターを使っても“恐らく”破れない”としか言えない）。

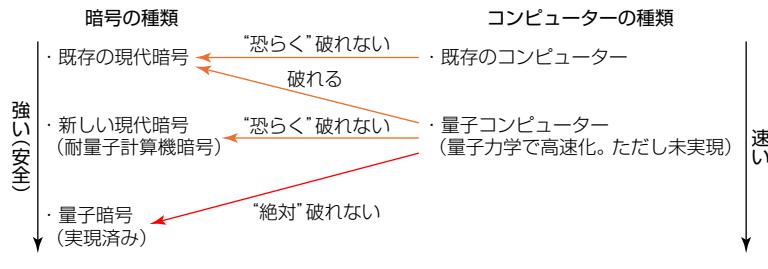


図2-各種の暗号と各種コンピューターの関係

一方、量子暗号方式の場合は、予想を一切導入せずに、最後まで厳密に証明できる。量子力学は、万有引力や電磁気の法則と同じく厳密に検証された物理法則である。その物理法則を前提にして、安全性が数学的に厳密に証明されている。したがって、将来いかなる解読法や新型コンピューターが出現しようとも解読されない。つまり“絶対に破れない”。

1.2 量子コンピューターとの比較

量子コンピューターも量子暗号も、どちらも、量子力学を使った情報技術（狭義には量子情報技術）である。

1.2.1 機能が全く異なる

その一方、差異も複数あり、まず機能が全く異なる（図2）。量子コンピューターはコンピューターであり、高速に計算することが目的である。一方、量子暗号の目的は、この章の冒頭で述べたとおり暗号であり、計算ではない。

1.2.2 実装難易度は量子暗号の方がはるかに低い

差異のもう一つは実装難易度である。有用な量子コンピューターは未実現だが、量子暗号は30年前に実現済みである。

量子コンピューターの理論は30年以上前からあり、その理論どおりに動く装置がもしあれば、計算が高速化できることも分かっている。しかしそのような装置の実装は困難であり、いまだに誰も成功していない。結果として、何らかの役に立つ計算を、既存のコンピューターよりも高速にやってみせたという報告はいまだにない。

それに比べると量子暗号の実装ははるかに容易である。誤解を恐れず単純化するなら、既存の光通信システムを微修正（送信側に減衰器を挿入し、受信側の検出器を高感度化）すれば量子暗号になる。そのため、1990年代には既に実装の報告が複数あり、暗号通信ができていた。続いて2000年代に複数のベンチャー企業が量子暗号装置を販売し始め、2020年前後には日本電気（株）や（株）東芝が事業化を開始した。

2. 当社の量子暗号への取組みと研究成果

当社は2000年前後から継続して量子暗号の研究に取り組んで、様々な成果を上げてきた。当初は装置実装と理論の両面で研究を行い、装置実装での大きな成果として“東京QKDネットワーク”への参画⁽²⁾があった。しかし2012年に実験研究から撤退し、それ以降は理論研究だけに注力して今日に至る。そこでこの章では、理論研究の成果だけを幾つか選んで述べる。なお当社の研究紹介ページ⁽³⁾で、更に詳しい情報を随時更新しているので、そちらも参照されたい。

2.1 理論研究をする意義

理論研究というと、実験結果を後追いで述べるだけと思われがちだが、量子暗号の場合は必ずしもそれにとどまらず、現実の量子暗号装置の性能向上やコスト削減に直接役立つ。そのことを一つの例で述べる。

2.1.1 成果例その1：スカッシュ演算子

2000年代前半までは量子暗号の理論が未発達であったため、量子暗号を作るには、単一光子を厳密に制御できる装置が必要と考えられていた(図3(a))。つまり、光源は光子を1回当たり厳密に1個発して、誤って0個、又は2個以上発してはならないとされていた。同様に検出器も、光子一つ一つを別々に検出できるとされていた。その前提の下、1.1.4項の前半で述べた意味で、安全性を証明していた。

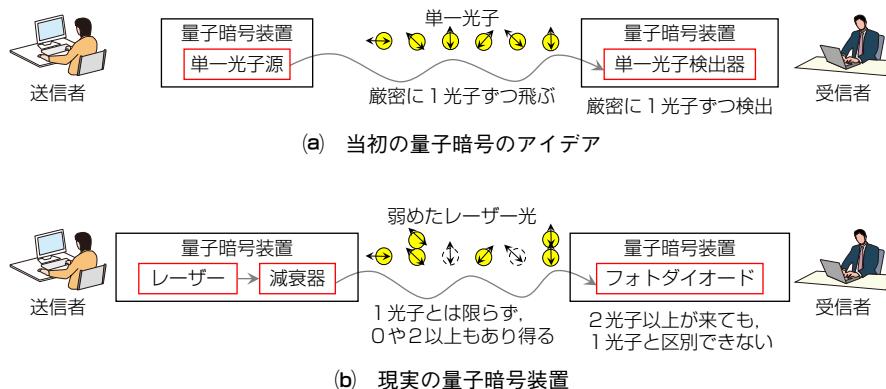


図3-当初の量子暗号のアイデアと現実の量子暗号装置

だが実は、当時も今も、そのような光源も検出器も存在しない。あっても近似的なものであり、かつ高価である。そこで実験家はそれらを安価な部品で代用することを考えて、光源にはレーザーを弱めたもの、検出器にはフォトダイオードを使用した(図3(b))。そうして作成した量子暗号装置はその場しのぎの廉価版と考えられて、実際に盗聴法も知られていた。だがその後理論研究が進展した結果、実はその廉価版の装置を用いても、厳密な安全性を達成できることが分かった。具体的には、まず盗聴法に対しては、“デコイ法”という新たなプロトコル(装置の操作手順)を採用すれば対処できると分かった。また検出器の問題は、“スカッシュ演算子”という数学的手法⁽⁴⁾を用いれば、厳密な安全性証明を書けるようになって解決すると判明した。なお後者のスカッシュ演算子は、当社とNTT(株)による共同研究成果である⁽⁴⁾。

2.1.2 理論研究でもたらされる装置の性能向上とコスト削減

2.1.1項の例では、かつてその場しのぎの廉価版とされていた量子暗号装置を、理論研究の成果によって、正規の装置に昇格できた。結果、理論研究によって、装置のコストを大幅に削減できたことになる。同様に、理論研究によって装置の性能を向上できる事例も多くある(2.2.1項、2.2.2項参照)。これが理論研究をする意義である。

2.2 その他の成果例

2.2.1 双対ユニバーサルハッシュ関数の提案

これは理論研究によって装置の性能が向上できた事例であり、名古屋大学との共同研究である。量子暗号の安全性を保証するには、秘匿性増強と呼ばれるデータ処理が不可欠である。そのデータ処理用の新たなアルゴリズムとして、当社は“双対ユニバーサルハッシュ関数”を提案した⁽⁵⁾⁽⁶⁾。このアルゴリズムのメリットは多々あるが、一例として、誤り率が小さい状況で最大の鍵生成速度を達成できることが挙げられる。

2.2.2 有限長解析による鍵生成速度の改良

これも2.1.2項と同様に、理論研究によって装置の性能が向上できた事例であり、名古屋大学との共同研究もある。Bennett-Brassard 1984(BB84)方式は最初に提案された量子暗号方式であり、現在でもデファクト標準になっている。当社はこのBB84方式に対する有限長解析(秘密鍵の長さが現実には有限であることを考慮に入れた安全性証明)を厳密に実施し、既存成果をしおぐ鍵生成速度の算出に成功した⁽⁷⁾。特に秘密鍵長が短い領域で、鍵生成速度が倍増される。

2.2.3 DPSQKD方式に対する最強攻撃と厳密スケーリングの導出

BB84方式以外にも様々な量子暗号方式が提案されている。とりわけ国産で有名なのが、2002年に日本電信電話(株)(現NTT(株))が提案したDPSQKD(Differential Phase Shift Quantum Key Distribution)方式である。当社は2007年にこの方式に対する攻撃手法を提案し、現在もこれが同方式に対する最強の攻撃手法だとされている⁽⁸⁾。

また2024年の富山大学との共同研究では、先に述べた攻撃手法を援用することによって、この方式の性能指標の一つである“スケーリング”の厳密な導出に成功し、それがBB84方式と同等であると示すことができた⁽⁹⁾。

2.2.4 異なる安全性証明手法を数学的に統合

この項では安全性証明の手法に関する研究成果について述べる。安全性証明では“Leftover Hashing Lemma(LHL)の手法”，及び“位相誤り訂正(PEC)による手法”という2種類の異なる数学的手法が広く用いられている。しかし2020年頃までは、これら両手法の関係は不明であり、両者は独立した手法とみなされていた。実際に、同一の量子暗号方式を、異なる手法で解析した論文が、それぞれ新たな成果として発表されることが常であった。これに対して当社は、両手法が実は、数学的には同一である、と厳密に示すことに成功した⁽¹⁰⁾⁽¹¹⁾。

2.2.5 放射線乱数の安全性を初めて厳密に証明

当社は、量子暗号の理論研究で培った知見を、以前からあった放射線を使った乱数生成器(以下“放射線乱数”という。)に適用し、東京大学工学部、大学共同利用機関法人 高エネルギー加速器研究機構(KEK)との共同研究として、放射線乱数のランダム性と秘匿性を、世界で初めて^(注1)理論的に厳密に証明することに成功した⁽¹²⁾⁽¹³⁾。なお証明に当たっては、原子核の持つパリティ対称性という性質と量子力学を活用した。詳細は参考文献⁽¹²⁾⁽¹³⁾及び東京大学工学部、KEKのプレスリリースを参照されたい。

(注1) 2022年7月7日、当社調べ

3. む す び

量子暗号とは何かを述べたのち、当社の理論研究の成果を述べた。量子暗号の場合は、理論研究の成果によって、現実の装置の性能向上やコスト削減が可能になるということを、幾つかの成果例とともに述べた。なお当社の研究紹介ページ⁽³⁾で最新の研究成果を随時更新しているので、興味のある方はそちらも参照されたい。

参 考 文 献

- (1) 鶴丸豊広：量子暗号、映像情報メディア学会誌、**69**, No.11, 889～897 (2015)
- (2) Sasaki, M., et al. : Field test of quantum key distribution in the Tokyo QKD Network, Optics Express, **19**, No.11, 10387～10409 (2011)
- (3) 三菱電機：三菱電機の量子技術
<https://www.mitsubishielectric.co.jp/corporate/randd/quantum/>
- (4) Tsurumaru, T., et al. : Security proof for quantum-key-distribution systems with threshold detectors, Physical Review A, **78**, No.3, 032302 (2008)
- (5) Tsurumaru, T., et al. : Dual Universality of Hash Functions and Its Applications to Quantum Cryptography, IEEE Transactions on Information Theory, **59**, No.7, 4700～4717 (2013)
- (6) Hayashi, M., et al. : More Efficient Privacy Amplification With Less Random Seeds via Dual Universal Hash Function, IEEE Transactions on Information Theory, **62**, No.4, 2213～2232 (2016)
- (7) Hayashi, M., et al. : Concise and tight security analysis of the Bennett-Brassard 1984 protocol with finite key lengths, New Journal of Physics, **14**, 093014 (2012)
- (8) Tsurumaru, T. : Sequential attack with intensity modulation on the differential-phase-shift quantum-key-distribution protocol, Physical Review A, **75**, No.6, 062319 (2007)
- (9) Mizutani, A., et al. : Tight scaling of key rate for differential-phase-shift quantum key distribution, Phys. Rev. Research, **6**, No.4, 043300 (2024)
- (10) Tsurumaru, T. : Equivalence of Three Classical Algorithms With Quantum Side Information: Privacy Amplification, Error Correction, and Data Compression, IEEE Transactions on Information Theory, **68**, No.2, 1016～1031 (2022)
- (11) Tsurumaru, T. : Leftover Hashing From Quantum Error Correction: Unifying the Two Approaches to the Security Proof of Quantum Key Distribution, IEEE Transactions on Information Theory, **66**, No.6, 3465～3484 (2020)
- (12) Tsurumaru, T., et al. : Secure random number generation from parity symmetric radiations, Communications Physics, **5**, No.147 (2022)
- (13) 暗号システム向けの理想的な乱数実現へ—“パリティ対称性”利用が有効—, 三菱電機技報, **98**, No.1, 1-3-01 (2024)

次世代データセンターの大容量化／低電力化を実現する光電融合技術

Opto-electronic Convergence Technology for Achieving High Capacity and Low Power Consumption in Next-generation Data Centers

白尾瑞基*
Mizuki Shirao
大畠伸夫*
Nobuo Ohata

*情報技術総合研究所(工博)

要旨

近年、LLM(Large Language Model)を代表とする生成AIの登場によって、マシンラーニングに必要な演算量が爆発的に増加し続けている。演算量の増加に伴って、AIデータセンターの電力も増加の一途をたどっており、大きな社会課題になっている。この課題に対する一つのソリューションとして、光電融合技術が注目を集めている。電気ICと光デバイスを小型に集積する光電コパッケージは、従来の光トランシーバーでは実現が難しい大容量通信と低消費電力化を実現でき、次世代のAIデータセンターでの導入が期待されている。

1. まえがき

LLMを代表とする生成AIの登場によって、マシンラーニングに必要な演算量が爆発的に増加し続けている。図1はプロセッサーの演算性能(FLOPS)、コンピューティングのネットワーク帯域、マシンラーニングで用いられているパラメーター数をそれぞれ規格化してプロットしたものである⁽¹⁾。マシンラーニングで使用されるパラメーター数は2年で約600倍のペースで増加しており、OpenAI社が開発したGPT-3^(注1)では機械学習で使用したパラメーター数は1,750億、Switch transformerやGPT-4^(注1)は1兆以上のパラメーターが使用されている。一方、プロセッサーの演算性能は2年で約2.6倍、コンピューティングに必要なネットワーク帯域も2年で約1.5倍であり、パラメーター数の増加に対してコンピューティングの性能向上は緩やかである。このため、数千を超える多くのGPU(Graphics Processing Unit)を光通信によって並列接続し、クラスター化することで、演算性能を大幅に向上させて、機械学習を実行している。しかし、演算には多くのプロセッサーや光通信機器が必要になり、AIデータセンターでは膨大な電力が必要になっている。

三菱電機はこの課題に対して、光通信の電力効率とネットワーク帯域拡大を両立する光電コパッケージの検討を開始している。本稿では、機械学習に必要なバックエンドネットワークのトレンドや光電コパッケージ形態の技術トレンドを示す。また、これまで開発してきた高速のEML(Electro-absorption Modulator integrated Laser)の特性について述べた後、光電コパッケージへの適用に向けた高集積技術について述べる。

(注1) GPT-3、GPT-4は、OpenAI OpCo, LLCの登録商標である。

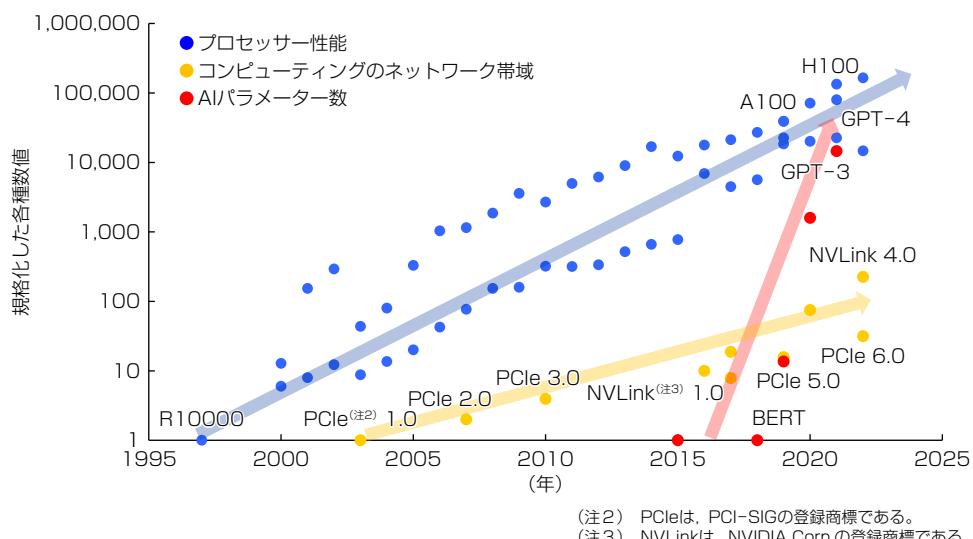


図1-プロセッサー性能、コンピューティングネットワーク帯域、AIパラメーター数のトレンド

2. AIマシンラーニングのネットワークトレンドと光電コパッケージの導入

図2にバックエンドネットワークの概念図を示す。複数のスイッチを介することで多くのGPUを並列に接続する構成を取っている。NVIDIA社ではNVスイッチを介して72台のGPUを、NVLinkと呼ばれる独自の高速ネットワークで並列接続しているが⁽²⁾、今後、演算能力の向上に向けて既存のGPU72台から576台、1,152台にスケールアップが進むと考えられる。GPUやスイッチ間を光通信で接続するための光トランシーバーの数量も増加すると見られており、消費電力が大幅に増加するという課題がある。また、GPUとの接続数が増えることでスイッチASIC(Application-Specific Integrated Circuit)の大容量化も進んで、スイッチASICから出力される信号帯域も合わせて増加する。そのため、光トランシーバーの通信帯域も広帯域化が必要になり、単位長さ当たりどれくらいの信号が取り出せるか、つまり帯域エッジ密度(Tbps/mm)が今後重要な指標になる。図3は、図1に示すトレンドを参考に、帯域エッジ密度を計算した結果である。帯域エッジ密度は2028年には1Tbps/mmが必要になり、2030年には2Tbps/mmと指数関数的に増大する。

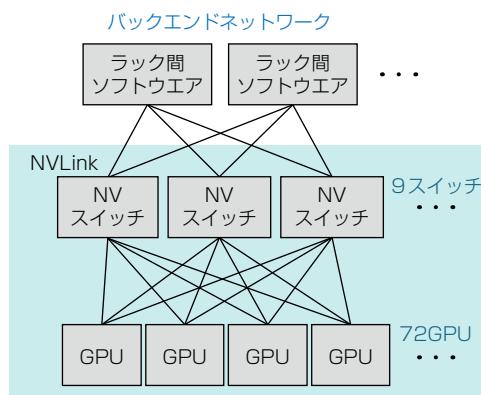


図2-バックエンドネットワークの概念図

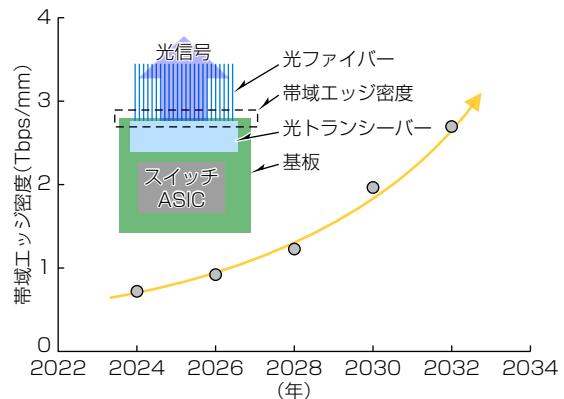


図3-帯域エッジ密度のトレンド

既存のプラガブル光トランシーバーでは低電力かつ高いエッジ密度を実現することが難しいため、これらの課題を解決する新技術として光電コパッケージが期待されている。表1にプラガブル光トランシーバーと光電コパッケージの比較を示す。プラガブル光トランシーバーはプリント基板上に実装された電気コネクターに接続して使用する構成である。そのため、GPUやスイッチASIC等のロジックICからトランシーバーまで距離が遠く、劣化した電気信号を補正するDSP(Digital Signal Processor)が必要になり、消費電力も大きい。電力効率はおおむね20pj/bitである。また、帯域エッジ

表1-プラガブル光トランシーバーと光電コパッケージの比較

	実装図	トランシーバー上面図	DSP 有無	電力効率 (pj/bit)	帯域エッジ密度 (Tbps/mm)
プラガブル 光トランシーバー			有	約20	<0.1
光電 コパッケージ			無	約5	<2

HBM : High Bandwidth Memory, PIC : Photonic Integrated Circuit

密度は、コネクターによる配線ピッチの制約もあり0.1Tbps/mm以下になる。一方、光電コパッケージはロジックICが搭載されるパッケージ基板上に実装されるため、DSPが不要になり、電力効率は5pj/bitが可能になる。また、微細配線可能な基板上に実装する形態であることから、帯域エッジ密度も2Tbps/mm以下が可能である。さらに将来はSi(シリコン)インターポーラー上に実装する形態も考えられて、その場合、電気のラインアンドスペース上は2Tbps/mmを超える帯域エッジ密度が可能になる。

当社はEMLを搭載する光電コパッケージの検討を開始した。EMLはDFB-LD(Distributed Feedback Laser Diode)と、電圧印加によって光を吸収するEAM(Electro Absorption Modulator)をワンチップに集積したレーザー素子である。高密度集積に優れるSiフォトニクス変調器よりも低電力で高速動作が可能であり、更に高い帯域エッジ密度の実現が期待できる。しかしながら、光電コパッケージの場合EMLを高密度に配置する必要があり、高周波配線とEMLをワイヤで接続する従来実装技術では実現困難である。そこで、フリップチップ実装を採用することを検討した。

3. 低電力、高エッジ密度実現に向けたEML実装技術

この章では、当社が開発した超高速EML⁽³⁾⁽⁴⁾について述べて、光電コパッケージへの適用可能なEMLの実装形態とその特性について述べる。

3.1 実装形態

当社EAMは、光吸収層の両側を低屈折率材料で挟んだハイメサ導波路構造を採用した独自構造を持つ(図4)。この構造によって、屈折率の高い光吸収層に光が強く閉じ込められるようになるため、効率的に光を吸収することが可能になる。言い換えると、静電容量の小さなEAMでも十分な光の吸収が得られるため、高速動作が実現できる。EAMはEMLの前方(光信号が出力される側)にあるため、EMLの横側に高周波配線を配置し、金ワイヤを用いてEAM電極へとワイヤボンディングを行うのが一般的である(図5)。ワイヤボンディングは広い実装面積が必要なため、過去の開発ではEML当たり1.25mm程度の幅を占有し、帯域エッジ密度を制約していた。今回、光電コパッケージに求められる高い帯域エッジ密度を達成するため、ワイヤボンディングを用いない実装技術を開発した。

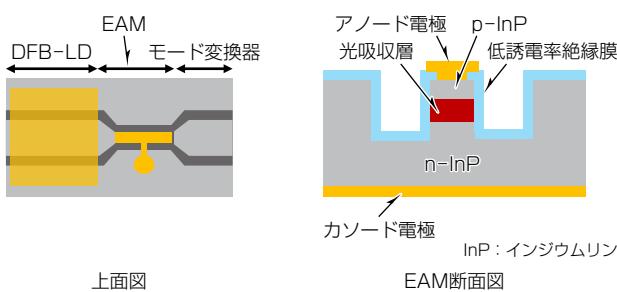


図4-EAMにハイメサ構造を採用したEML

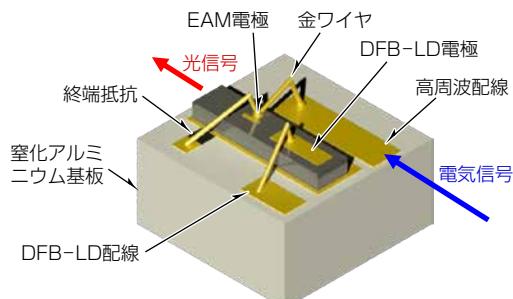


図5-ワイヤボンディングを用いた従来の実装形態

3.2 特性評価

光電コパッケージの形態を図6に示す。光電コパッケージでは、電気ICと光デバイスがフリップチップ技術を用いて実装される。フリップチップ実装では、電極同士をバンプ等によって直接接合するため、図5に示したようなワイヤボンディングを使用しない。これによって、従来の構造に比べて大幅な実装密度の向上が可能になるだけでなく、ワイヤボンディングによる寄生インダクタンス成分がなくなるため、動作速度の向上が期待できる。フリップチップ実装でのシミュレーションでは、145GHzの動作速度が示されている(図7)。

実験的には、窒化アルミニウム基板上へのフリップチップ実装検証によって、90GHzの動作速度を達成した⁽⁵⁾。小信号特性を図8に示す。小信号特性には図7では見られない凹凸が観測されており、これは高周波配線でのインピーダンス不整合に起因するものと考えられる。今後、製造技術の進展によってEMLのアレー化が達成されれば、従来の5倍になる0.25mm程度まで高密度化が期待できる。

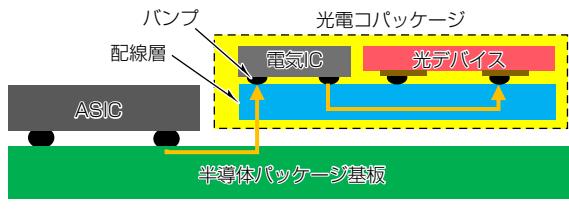


図6-光電コパッケージでの実装形態

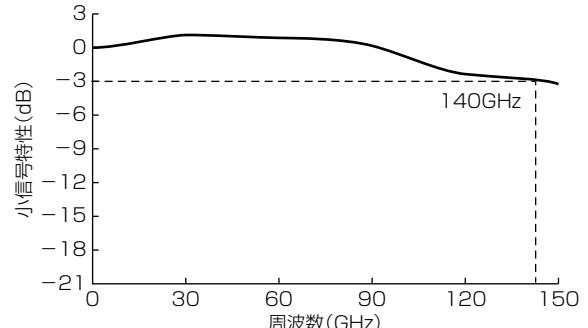


図7-光電コパッケージの動作帯域シミュレーション結果

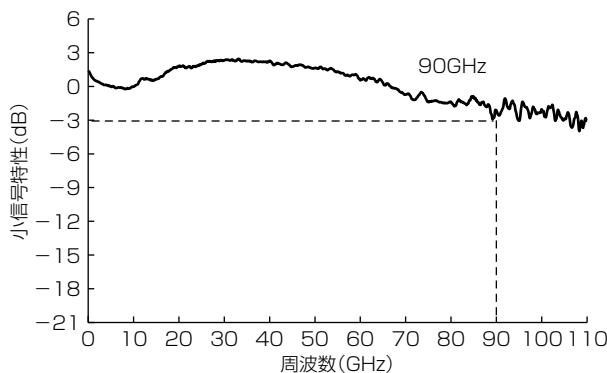


図8-光電コパッケージを想定した実装でのEMLの小信号特性(5)

3.3 今後の展望

次世代データセンターの大容量化に向けた今後の展望として、動作速度の改善に向けたEAM静電容量低減やアレー化を含む素子構造の最適化が必要である。また、低誘電率な基板材料(例：石英基板)を用いることで、更なる動作速度の向上が期待される。これらを達成するとともに、次世代の変調方式である200GBaud PAM4(Pulse - Amplitude - Modulation 4 - Levels)を適用することで、1.6Tbps/mmのエッジ密度達成が見込まれる。

4. む す び

バックエンドネットワークと光電コパッケージのトレンドを示すとともに、EMLの特性と光電コパッケージに向けた取組みについて述べた。ハイメサ構造を持つEMLを用いたフリップチップ実装技術によって、光電コパッケージに求められる高い帯域エッジ密度が実現できる。今後、AIマシンラーニングが進展する中で、光電融合技術の重要性は更に増していくと見られており、光デバイス技術の進化に加えて、光エレクトロニクスの実装技術の進化が必要不可欠になる。引き続き次世代の光電融合技術に向けた開発に取り組んでいく。

参 考 文 献

- (1) riselab : AI and Memory Wall
<https://medium.com/riselab/ai-and-memory-wall-2cb4265cb0b8>
- (2) NVIDIA : NVLinkとNVSwitch : 最速のHPCデータセンタープラットフォーム
<https://www.nvidia.com/ja-jp/data-center/nvlink/>
- (3) Uchiyama, A., et al. : Demonstration of 155-Gbaud PAM4 and PAM6 Using a Narrow High-Mesa Electro-Absorption Modulator Integrated Laser for 400 Gb/s Per Lane Transmission, Journal of Lightwave Technology, **43**, No.4, 1868~1873 (2025)
- (4) Shirao, M., et al. , High Speed EML and Assembly Techniques for GPU Cluster System, 2025 Optical Fiber Communications Conference and Exhibition, 1~3 (2025)
- (5) Masuyama, K., et al. , EML Assembled with Flip-Chip Technology on AlN Sub-mount Operating at 212.5 Gbps PAM4, OECC 2025, WG2-2 (2025)

三菱電機株式会社