Keywords you should know. 第157回

知っておきたいキーワード

MLOps
～ML技術の発展とMLOpsの変遷～

山本祐也†

† Weights & Biases Japan

"MLOps: The Evolution of Machine Learning Technologies and MLOps Practices" by Yuya Yamamoto (Weights & Biases Japan, Tokyo)

キーワード：Machine Learning，Deep Learning，MLOps，LLMOps，WandB，W&B

まえがき

MLOpsという用語が一般的に広まってから5年以上が経ちました．MLOpsとは何かというと，機械学習の開発，導入，および運用に関するベストプラクティスを集約したものです．近年のAI技術の進歩は実に目覚ましいものですが，それらに関するベストプラクティスという性質上，MLOpsの概念も並行して進化してきました．MLOpsに含まれる要素は今後も変わっていく可能性がありますが，2024年初頭の現在では，例えば以下が含まれます．

・データ，実験管理
・モデルマネジメント＆デプロイメント
・パイプライン自動化/モデルCI（Continuous Integration）
・モデルモニタリング
・Continuous AI
・コラボレーション＆ガバナンス

本稿では，ML技術の急速な進化とそれに伴うMLOpsの発展について，MLOps製品を提供するソフトウェア企業の立場から振り返り，今後のMLOpsの方向性について考察します．

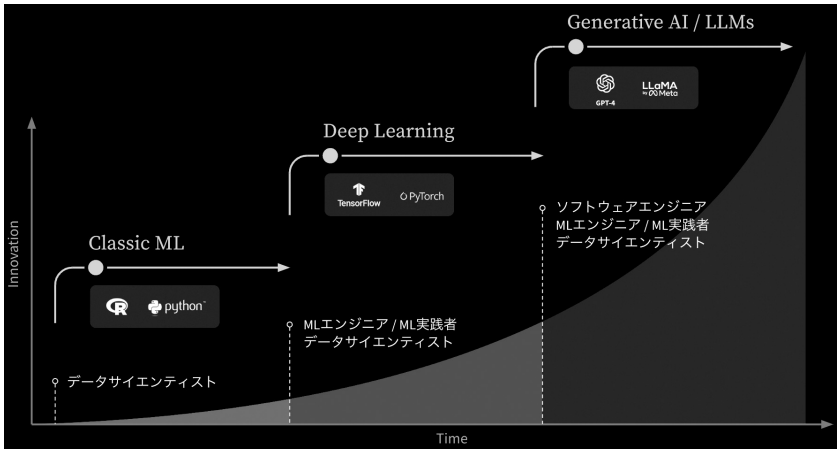


図1　機械学習技術の変遷の大まかな流れ

2010年代前半：Traditional MLの時代

2012年の「Harvard Business Review」誌の有名な記事「データサイエンティストは今後10年で最もセクシーな職業になる」は，AI・データサイエンスブームの火付け役となりました．当時の機械学習分野はデータサイエンティストが牽引し，彼らは統計学により近い存在でした．実際，データサイエンスという用語自体，統計学では学生を惹きつけられなかったアメリカの大学の事情から生まれたものです．

この時代の機械学習は，主に手作業でデータから特徴量を抽出し，モデルを訓練する手法が中心で，ビジネス現場での表形式データに対して有効でした．データの前処理や特徴量エンジニアリングに多くの手作業を要したため，このプロセスを自動化するAutoML製品が登場し，私の前職で取り扱っていたDataRobotなどは先駆的な例でした． ☞

☞ 2010年代前半の終わりごろ，機械学習技術の産業界への普及が進み，MLOpsに対するニーズが生まれ始めました．この「MLOps」という用語は，ParallelM社がDevOpsになぞらえて提唱したものです．当時のMLOpsのアプローチは，モデルの完成後の管理や監視に主眼を置いており，ParallelMのプラットフォームも同様でした．これは，ディープラーニング以前の伝統的な機械学習では，特徴量エンジニアリングにセンスこそ必要だったものの，モデルの訓練自体はシンプルで時間もかからなかったためです．そのため，「モデリングはそれほど重要ではなく，ビジネス理解や運用の方が重要」という声が現在より強く，そうした意見がSNS上でよく喝采を集めていました．この状況は後にディープラーニングの登場により変化することになります．

続いて，2010年代後半のディープラーニングの時代の到来と，産業界へのその影響について詳しく見ていきましょう．Hinton教授らのILSVRC2012での記録的な成果が，ディープラーニングの時代のスタートを切りましたが，産業界でのその広範な普及はもう少し後の話となるためです．

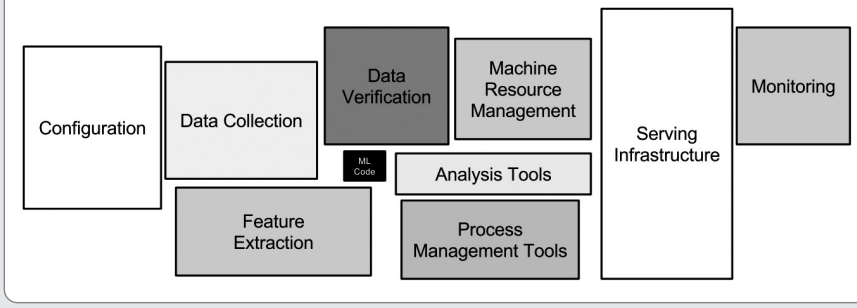


図2　MLシステム全体においてML自体はコード量も重要性もしばしば小さな要素とされがちだった（出典：Hidden Technical Debt in Machine Learning Systems, 2015）

2010年代後半：Deep Learningの時代

2010年代後半，以前は特別視されることもあったディープラーニングが産業界に本格的に普及しました．2015年から2016年にかけて登場したTensorFlow，Chainer，PyTorchなどのフレームワークにより，多くのユーザがディープラーニングの技術を当たり前に身につけていきました．ディープラーニングは，特に画像認識や自然言語処理の分野で，従来の機械学習では解決できなかった多くの課題を解決し，現代社会のさまざまな場面で広く活用されています．

さて，MLOpsは機械学習のベストプラクティスを集約したものであり，機械学習自体の変化に応じて，そのあり方も変化します．従来の機械学習とディープラーニングでは，取り組むべき課題や問題点が異なります．ディープラーニングの主な課題には，「訓練プロセスの複雑さ」や「非構造化データの取り扱い」などがあり，大量の計算リソースを使用し，長時間の訓練が必要な場合もあります．このような状況では，モデリング自体がもはや単純な作業ではなくなり，ここにもベストプラクティスが求められるのです．

Weights & Biases社の製品は，ディープラーニングネイティブなMLOps製品として開発され，新しい時代の要求に応えるための機能を備えています．具体的には，実験管理，ハイパーパラメータ最適化，可視化やコラボレーション機能などがあります．これらの機能により，ディープラーニングの開発プロセスがより効率的で再現可能になり，チーム間での共同作業による生産性向上も促進されます．

MLOpsの観点から，ディープラーニングのトレーニングにおいて重要なのはまず実験管理です．これには，モデルの学習進行状況をリアルタイムでモニタリングすること，そして実験の条件を自動的に記録し，再現性を高めることが含まれます．これらの情報を管理することで，実験の結果が信頼性のあるものになり，再現性が担保されます．

図3の例では，自動運転に関連するBerkeley DeepDrive（BDD）データデータを用いてセグメンテーションモデルのトレーニングを行った際のモデル精度を示しています．グラフの各ラインは，ディープラーニングの実験の一連のプロセスを追跡する軌跡となっています．これにより，異なる実験条件下でのモデルの振る舞いや性能の変化を視覚的に比較することができ，最適なモデル構成やパラメータを ☞

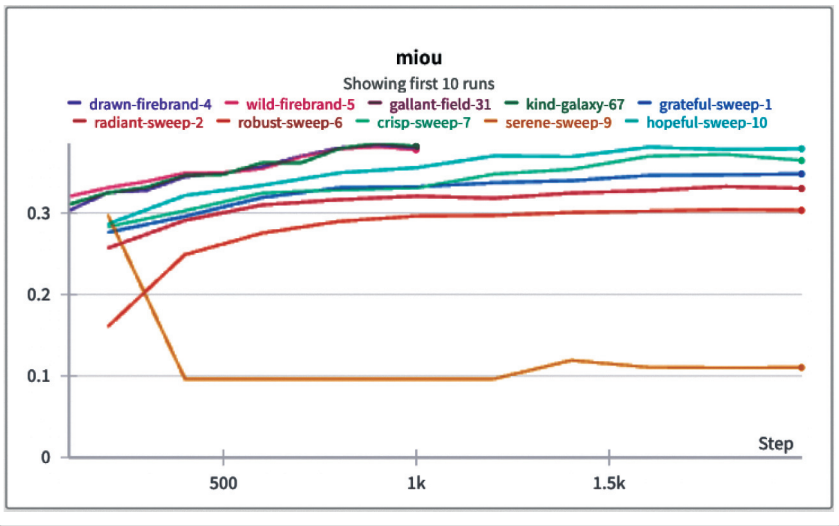


図3　セグメンテーションタスクにおける平均IoUスコアのモニタリングの例
（IoU: Intersection over Union，正解と予測の重なりの良さを表す指標）

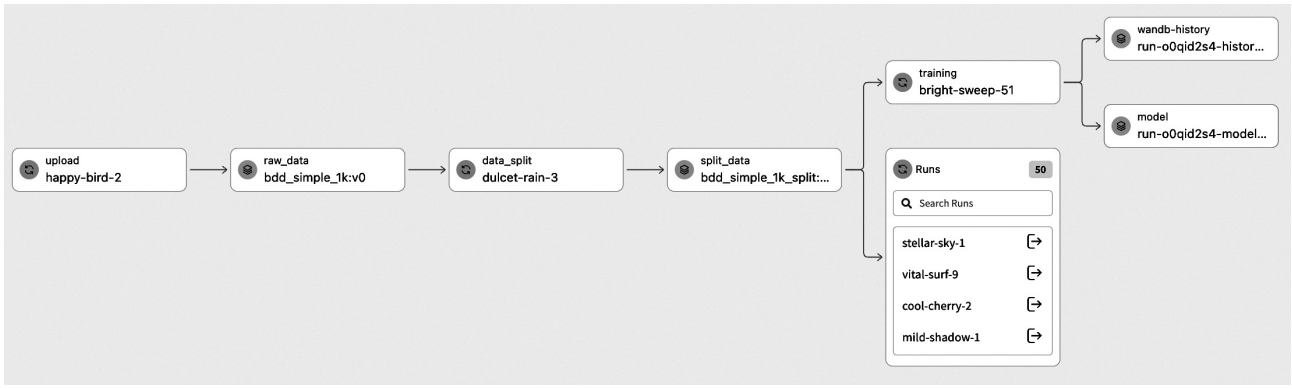


図5　ディープラーニングの一連の実験の系譜を表すリネージ

図4　WandBのイメージオーバレイ機能によるセグメンテーション（左）および物体検出（右）の可視化と分析

☞ 選択できるようになります．

機械学習プロジェクトでは，実験結果の分析と洞察に基づくモデルの改善が重要です．特にディープラーニングでは，非構造化データを扱うため，このプロセスはより複雑になります．実験の条件やパフォーマンス指標を非構造化データと関連付けて分析し，モデルの問題点を明確にし，改善策を策定する必要があります．

画像のセグメンテーションや物体検出タスクにおいて，モデルの予測結果と実際のデータを比較することで，特定のデータポイントや属性に関連する問題点を素早く見つけることができます（図4）．これは，開発を効率的に進める上で重要です．例えば，モデルが自転車の検出に問題があれば，自転車を含むデータを抽出し，予実差を確認することで，改善の手がかりが得られるでしょう．

ディープラーニングプロジェクトの複雑さを考慮すると，どのデータセットを使ってどのモデルが作成され，そのモデルからどのような推論結果が導かれるかを示す「家系図」の自動作成は，その透明性と追跡可能性を高める上で重要です（図5）．プロジェクトのリネージ（系譜）を自動追跡できれば，特定の結果がどのモデル，データセット，前処理スクリプトに由来するのかを常に一目で把握できます．

このようにディープラーニング時代のMLOpsは，モデルができた後だけでなく，モデル学習時の実験管理やインサイトの抽出，トレーサビリティの担保といった領域もカバーすることが求められています．Weights & Biases（WandB）はこれらのすべての機能を提供するこの分野の代表的な製品です．

ディープラーニング時代のMLOpsについて，WandBを例に取り上げながらいくつかの要素を紹介してきましたが，これらはいずれも伝統的な機械学習の時代には含まれなかった構成要素です．現代のディープラーニングは，これらの要素が連携し，スケーラブルに自動化されることで支えられています．

ここで余談ですが，MLOpsという用語を生み出したParallelM社のその後について触れたいと思います．ParallelM社は2019年にDataRobot社に買収されました．ParallelM社はこの分野のパイオニアであり，MLOpsという概念を形作る上で重要な役割を果たしましたが，当時のMLOpsの認知度は非常に低く，ほぼ無名に近い状態でした．少し時代を先取りし過ぎたのかもしれませんね（図6）．これはトリビアですが，「MLOps」という用語はアメリカでParallelM社が商標登録していたという事実があり，現在ではDataRobot社がその商標を保有しています．

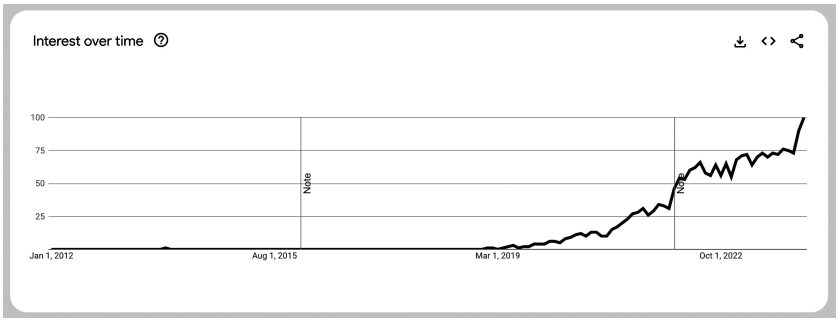


図6　「MLOps」に対するGoogle TrendsのInterest over time

2020年以降：Generative AIの時代

2020年代は，ChatGPTの公開をきっかけに，大規模言語モデル（LLM）を中心とする生成AIが急速に発展しています．LLMは，人間のような対話が可能で，膨大な知識と多様なタスクへの適用性を持っています．この時代の機械学習の特徴は，巨大な基盤モデルを大量のデータで事前学習し，それを多様なダウンストリームタスクに適応させるアプローチです．以前のディープラーニングでも似た手法はありましたが，基盤モデルでは特徴抽出能力だけでなく，具体的な知識の獲得も事前学習で行われ，その規模も格段に大きくなっています．また，APIを通じて大規模基盤モデルを利用することも一般的になりました．

事前学習の規模が大きくなるにつれ，実験管理はより重要性を増しており，OpenAIなど基盤モデルを開発する企業のほとんどがWeights & Biasesを用いています．さらに，基盤モデルの開発者とアプリ開発者が異なるケースも増え，プロンプトエンジニアリングという新しい役割も出現しました．プロンプトエンジニアは，生成AIを操作するための指示文であるプロンプトを巧みに設計し，ダウンストリームタスクの解決に貢献しています．

この新しい流れの中で，LLMOpsという用語が徐々に使われるようになりました．これは，LLMを中心とした新しい時代のMLOpsの形態を指しています．LLMOpsはまだ概念として成熟しているとは言えませんが，市場ではいくつかの重要な共通点が認識されています．例えば，プロンプトやその他のコンポーネントに関わるさまざまな設定や入出力データの追跡です．

この追跡機能は，「Retrieval Augmented Generation（RAG）」のような手法において特に重要です．RAGでは，入力クエリに基づいて関連情報をベクトルデータベースから取り出し，その情報をLLMへの入力として用いるコンテクストとして組み合わせます．このプロセスで問題が生じた場合，複数のコンポーネントを組み合わせた処理チェーンのどこに問題があるのかを迅速に特定する必要があります．そのため，各種コンフィグや入出力データの詳細な追跡は，問題の特定と解決を迅速に行う上で不可欠なのです．なお，図7は執筆時点のWeights & Biasesのプロンプトトレース画面ですが，本稿の公開までに大幅に刷新される予定であり，このことも機械学習技術の急速な進歩とMLOpsの変遷のスピード感を感じさせます．

本稿を通じて，2010年代以降の機械学習技術の急激な進化と，それに伴うMLOpsの発展について概説してきました．技術進歩の加速度が増す中，MLOpsの発展も同様に速いペースで進んでいくことには疑いの余地がありません．この変化の中で，次の時代を見据えるMLエンジニアの皆さまにとって，本稿が今後の取り組みや思索において，少しでも役立つ参考資料となれば幸いです．

（2024年3月19日受付）

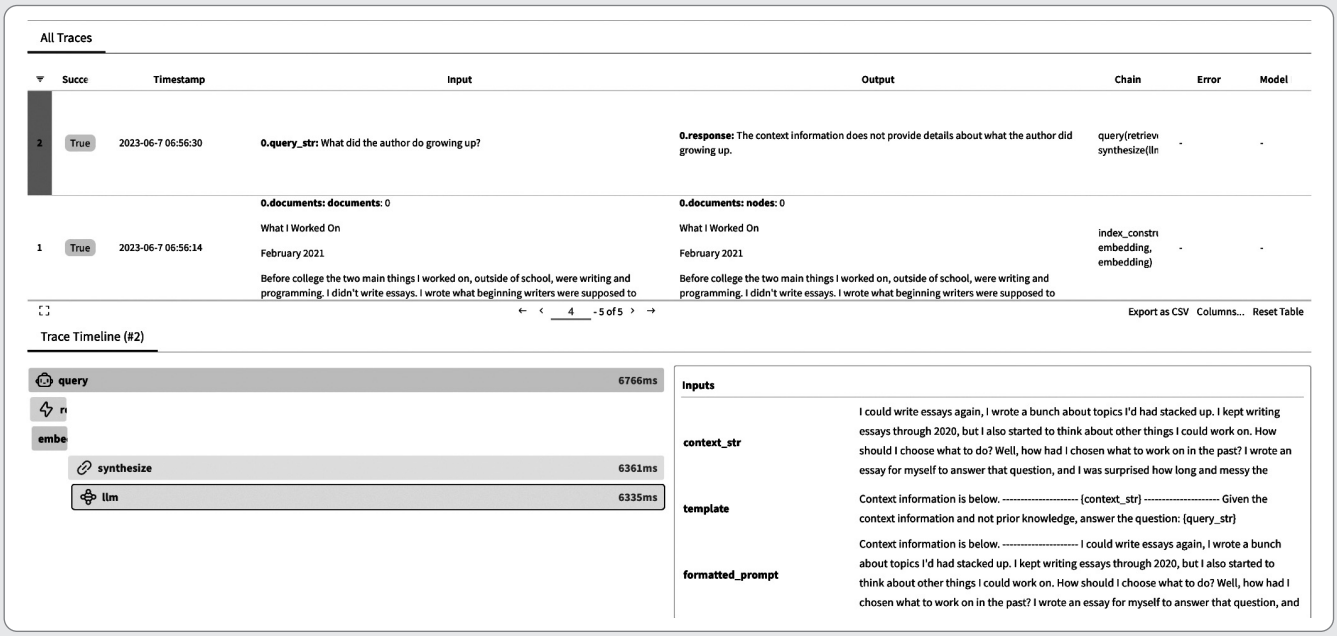
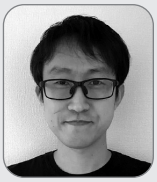


図7　WandB Tracesによるプロンプトのトラッキング例



山本　祐也（やまもと　ゆうや）　東京大学大学院で博士号取得後，大手製造業を経て，前職DataRobotで製造領域担当として国内のAI導入を支援．現在は，Weights & Biasesで機械学習エンジニアとして，WandBの導入支援やコミュニティ活性化に取り組む．Kaggle Grandmasterの一人．