

# 機械学習研修Day3

## (ハンズオン・事例紹介)

酒井 敏彦, 渡辺 龍二 / ペパボ研究所

2022.08.04 機械学習研修

GMOペパボ

1. 研修のおさらい(数分)
2. ハンズオン(105 min)
3. ペパボでのサービス連携事例紹介(50 min)
4. 学んだことを各自まとめる(15 min)

## 機械学習研修の目的はこれだ！

サービス価値の向上のために「再帰化」を実現する上で、実装手段として機械学習を検討できるようになる。

- 目的の背景:ペパボのエンジニアとして機械学習を学ぶ動機は、  
機械学習を「再帰化」の実現アプローチとして利用することにある
  - サービス価値の向上において「再帰化」は重要な考え方<sup>[栗林2022]</sup>
  - 「再帰化 ≠ 機械学習の利用」であることを理解して  
再帰化を実現するために機械学習の利用するか否かを判断できることを目指す

研修目的を達成するために以下のカリキュラムを設定しています！

研修の目的

サービス価値の向上のために「再帰化」を実現する上で、実装手段として機械学習を検討できるようになる。

	ゴール	実施内容
Day1	機械学習とは何か、 自分の言葉で説明できる	<ul style="list-style-type: none"> <li>- 機械学習イントロダクション(座学)</li> <li>- Machine Learning Crash Course (ハンズオン)</li> </ul>
Day2	再帰化を実現するために、 機械学習をどう活用できるか イメージできる	<ul style="list-style-type: none"> <li>- Machine Learning Crash Course (ハンズオン)</li> <li>- 再帰化と機械学習(座学)</li> </ul>
Day3	サービスの実課題を 機械学習で解決する 一連の流れの例を挙げられる	<ul style="list-style-type: none"> <li>- Machine Learning Crash Course (ハンズオン)</li> <li>- 機械学習のサービス連携事例紹介(座学)</li> </ul>

1. 研修のおさらい(数分)
- 2. ハンズオン(105 min)**
3. 機械学習とサービスの連携事例紹介(50 min)
4. 学んだことを各自まとめる(15 min)

## ハンズオンを引き続きやっていきましょう！

- [Machine Learning Crash Course](#)
- 今日進めて欲しい単元
  - Classification (75min)
    - Video Lectureと  
ROC Curve and AUCは  
飛ばして大丈夫です
- 時間が余ったときに進めて欲しい単元
  - Representation
  - Feature Crosses
  - Regularization: Simplicity

### Machine Learning Crash Course with TensorFlow APIs

Google's fast-paced, practical introduction to machine learning, featuring a series of lessons with video lectures, real-world case studies, and hands-on practice exercises.

[Start Crash Course](#)[View prerequisites](#)

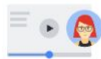
30+ exercises



25 lessons



15 hours



Lectures from Google  
researchers



Real-world case studies



Interactive visualizations

## 進める上での注意点

- ハンズオン資料は英語です
- はあ？どゆこと？となったらすぐに聞いてください！
- 動画は飛ばしてください！
- 休憩を適宜取ってください
- 早く終わった場合はどんどん先に進んで大丈夫です！

1. 研修のおさらい(数分)
2. ハンズオン(105 min)
- 3. 機械学習とサービスの連携事例紹介(50 min)**
4. 学んだことを各自まとめる(15 min)



# 機械学習とサービスの連携事例紹介

渡辺 龍二, 財津 大夏, 酒井 敏彦  
2022.08.04 機械学習研修 Day3

GMOペパボ

- A. イントロダクション
- B. 機械学習のサービスへの導入
- C. サービスにとって有用な「関数」についての検討

サービスと機械学習を連携するために、次のステップを踏む必要がある(※超大雑把)

1) サービスのビジネス的な背景から  
サービスにとって有益な「関数」を定める



2) 関数をデータから実際に獲得する



3) 関数をサービスに導入する

ここまでCrash courseでやってきたのは 2)の部分

1) サービスのビジネス的な背景から  
サービスにとって有益な「関数」を定める



2) 関数をデータから実際に獲得する

この理論的背景を  
しっかり学んできたのが  
ここまでの本研修！



3) 関数をサービスに導入する

Day3ではこの全体を知るために…

1) サービスのビジネス的な背景から  
サービスにとって有益な「関数」を定める



2) 関数をデータから実際に獲得する



3) 関数をサービスに導入する

今から以下の2つについて一緒に考えていく！

1) サービスのビジネス的な背景から  
サービスにとって有益な「関数」を定める



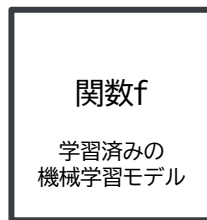
2) 関数をデータから実際に獲得する



3) 関数をサービスに導入する

- A. イン트로ダクション(by 渡辺)
- B. 機械学習のサービスへの導入**
  - a. イン트로ダクション(by 渡辺)
  - b. 実際の導入事例(by 財津)
- C. サービスにとって有用な「関数」についての検討
  - a. イン트로ダクション(by 渡辺)
  - b. ビジネス課題から考える機械学習の適用(by 酒井)

関数 $f$ (学習済みの機械学習のモデル)が既にあるとする

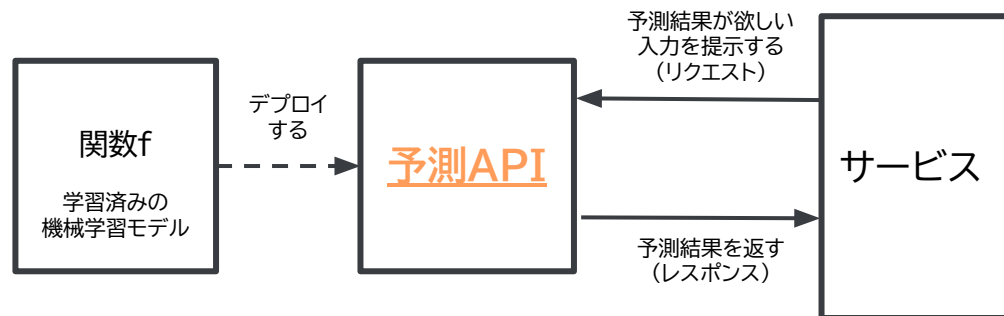




この関数 $f$ を「サービスと連携する」とは、どういうことを考えていく



連携の例: 予測APIを立て、必要な予測結果を逐次的に得るために通信する

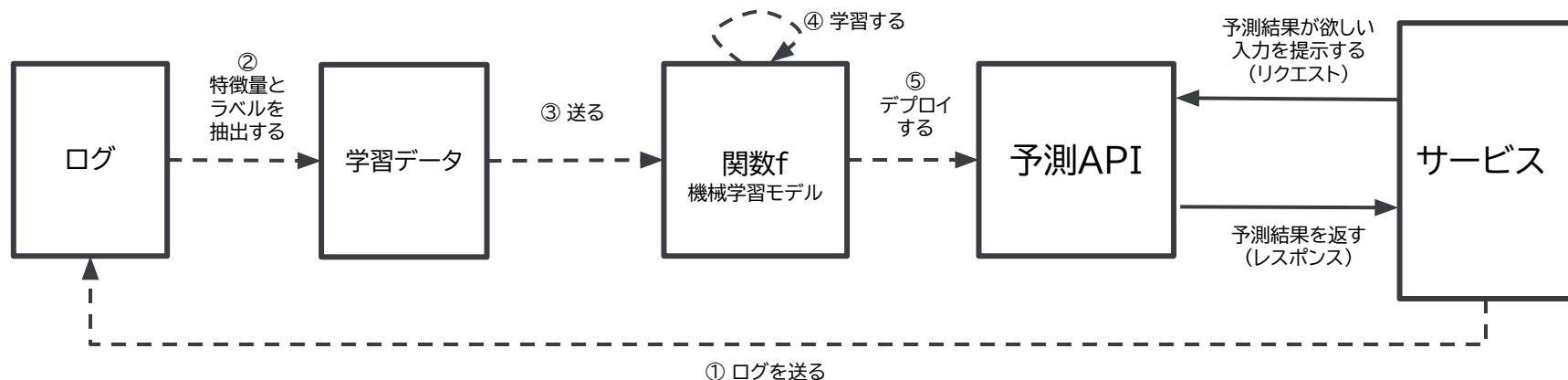


:オブジェクト

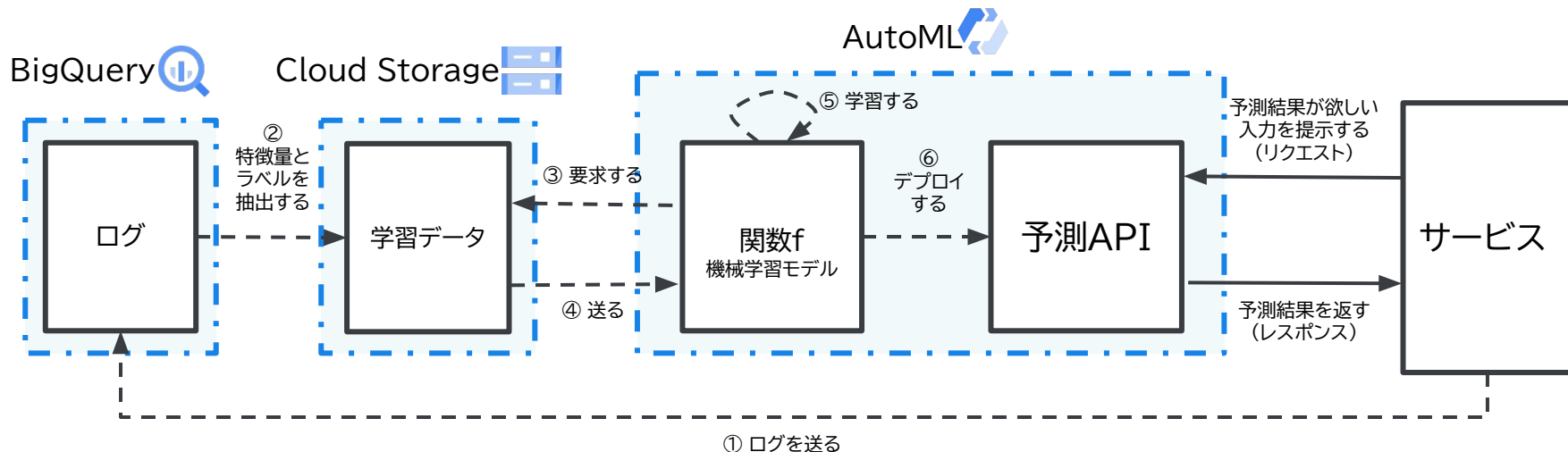
- - -> :一度しか実施しない処理

——> :繰り返し行われる処理

モデルが学習される過程まで含めると以下のようなになる(再帰化…！)



ペパボで利用している技術スタックによる1つの構築例は以下のようなになる



:利用するサービス

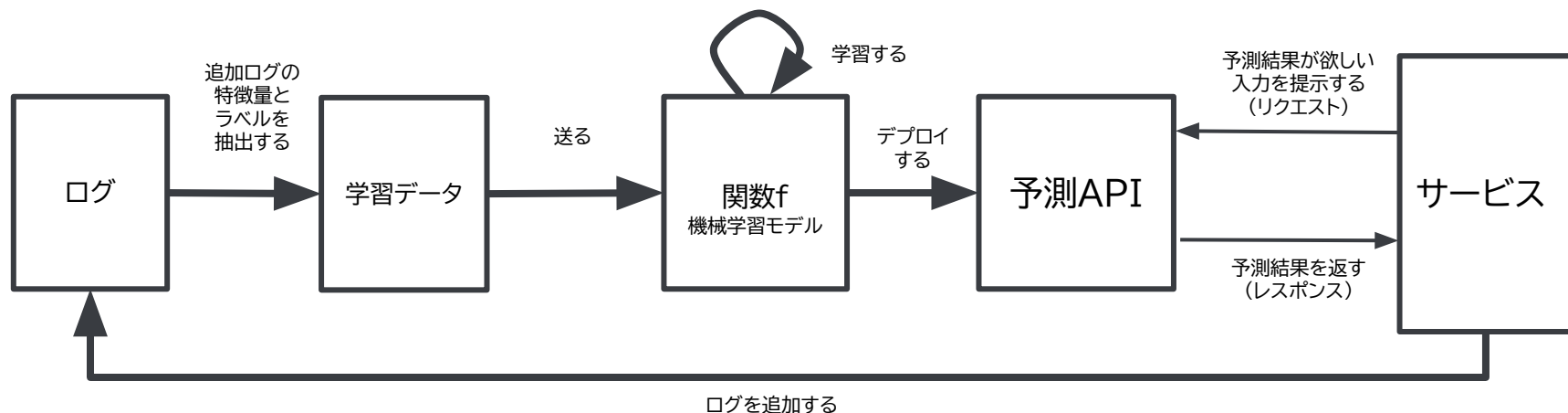
---▶:一度しか実施しない処理



:オブジェクト

—▶:繰り返し行われる処理

さらに、このループを繰り返し関数 $f$ を定期的に更新できるとより良さそう  
(もっと再帰化…！)



:オブジェクト

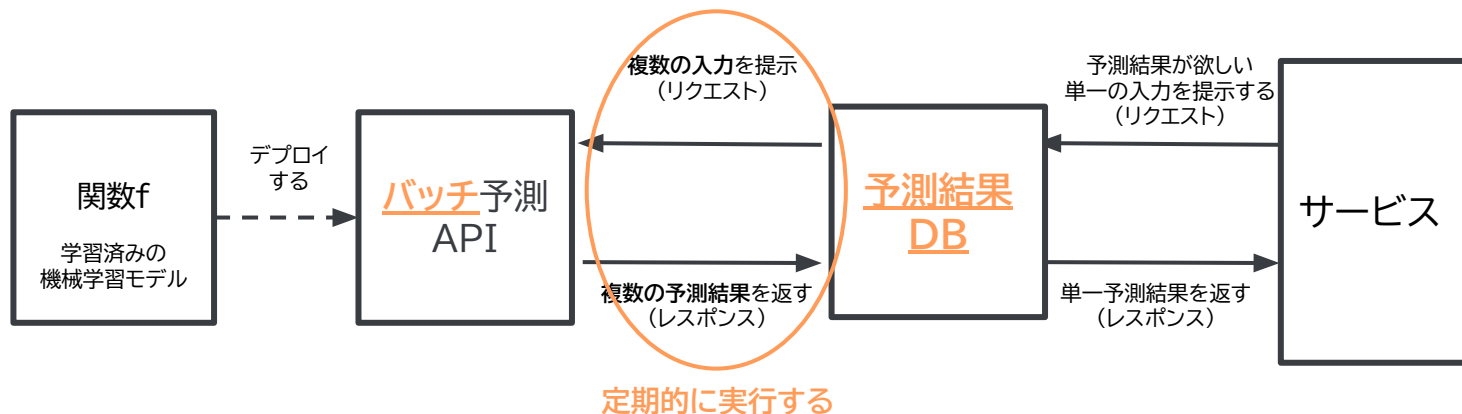


:繰り返し行われる処理

高速化のために、予測結果をバッファしておくようなことも考えられる(バッチ予測)

以下のような場合の対応として、バッチ予測が考えられる

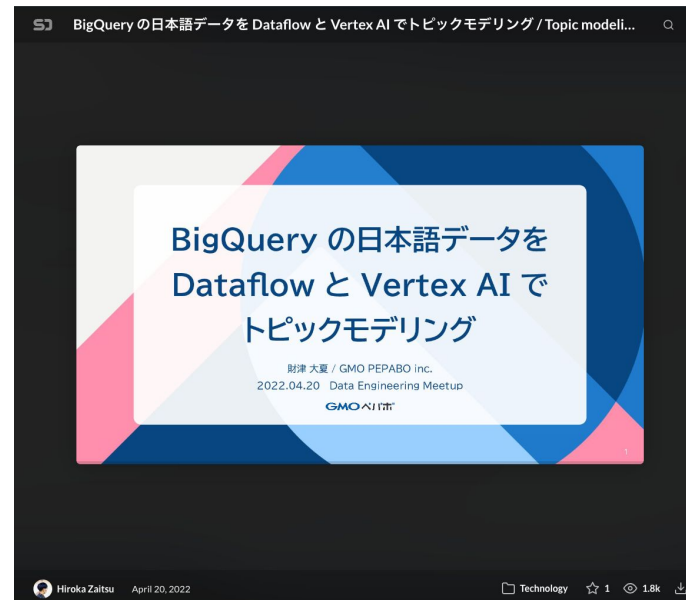
- 予測の対象や結果が一定期間変わらないと見なせ、都度予測するとコスパが悪い
- 予測に計算コストがかかり、レスポンスが低速



※その都度、入力を1つ受け取り予測結果を返すものはオンライン予測と呼ばれる

より詳しい機械学習のサービス導入事例を見ていきましょう！

- 発表タイトル: BigQuery の日本語データを  
Dataflow と Vertex AI でトピックモデリング
  - ペパボとZOZOの勉強会で発表された内容です



BigQuery の日本語データを Dataflow と Vertex AI でトピックモデリング  
<https://speakerdeck.com/zaimy/topic-modeling-of-japanese-data-in-bigquery-with-dataflow-and-vertex-ai>

- A. イン트로ダクション(by 渡辺)
- B. 機械学習のサービスへの導入
  - a. イン트로ダクション(by 渡辺)
  - b. 実際の導入事例(by 財津)
- C. サービスにとって有用な「関数」についての検討**
  - a. イン트로ダクション(by 渡辺)
  - b. ビジネス課題から考える機械学習の適用(by 酒井)



そもそも、機械学習で得る「関数」ってどんなものが良いんだろう…？

- そもそも、サービス価値向上のために「機械学習」を適用し再帰化したい
- そのために考えなければいけないことは何か？
  - サービスの課題を解決することに寄与する関数とはどういうものか？
    - 機械学習のターゲットとなる「関数」がサービスにとって無価値ならやる意味がない

実際の事例として酒井さんの取り組みを見ていきましょう！

- 発表トピック:minneの作品における「特徴ラベル」の自動付与の試み
  - 2022年度 情報処理学会関西支部 支部大会(9月に開催予定)にて発表予定の内容です  
※資料のオンライン公開は後日になります

1. 研修のおさらい(数分)
2. ハンズオン(105 min)
3. 機械学習とサービスの連携事例紹介(50 min)
- 4. 学んだことを各自まとめる(15 min)**

## 今日学んだことのまとめ・アウトプットの時間にします～

- Day3のゴール「サービスの実課題を機械学習で解決する一連の流れの例を挙げられる」を達成できそうでしょうか…？
- ぜひ以下のことをまとめてみてください～！
  - 学んだこと、難しかったこと、考えたこと、思ったことなど