

量子アニーリングマシンを活用したブラックボックス最適化の活用事例：
血行動態予測モデルのハイパーパラメータ探索

株式会社 Jij
牧野大介、宮脇悠、山城悠

中外製薬株式会社
田村彩、城取萌望、若宮慶也、中臣政司

1. 目的、背景
2. 問題設定
3. 手法の説明
4. 各種実験結果
5. まとめと今後の展望

目的・背景

最適化問題に対するアプローチ

数理最適化

目的関数と制約条件を数理的な手法で表現

Problem: TSP

$$\min \sum_{i=0}^{\text{len}(d,0)-1} \sum_{j=0}^{\text{len}(d,0)-1} \sum_{t=0}^{\text{len}(d,0)-1} d_{i,j} \cdot x_{t,i} \cdot x_{(t+1) \bmod \text{len}(d,0),j}$$

s.t.

$$\text{onehot_city} \quad \sum_{t=0}^{\text{len}(d,0)-1} x_{t,i} = 1 \quad \forall i \in \{0, \dots, \text{len}(d,0) - 1\}$$

$$\text{onehot_time} \quad \sum_{i=0}^{\text{len}(d,0)-1} x_{t,i} = 1 \quad \forall t \in \{0, \dots, \text{len}(d,0) - 1\}$$

where

x 2-dim binary variable

問題に適したソルバーで求解

- Gurobi
- CPLEX
- D-Wave Leap
- Digital Annealer

ブラックボックス最適化

目的関数をデータ駆動で表現

モデルベース

ベイズ最適化

FMA

サロゲートモデル

ガウス過程等

FM

獲得関数

EI, PI, UCB等

FMで近似された目的関数

最適化

勾配ベースの手法等

量子・イジングマシン

Factorization Machine

Factorization Machine with Annealing (FMA)

- モデルベースのブラックボックス最適化手法
- サロゲートモデルとしてFactorization Machine (FM)を用いる。

Factorization Machine (FM)の特徴

- 数理モデルがQUBOの一例と見做せるため、アニーリングマシンとの相性が良い。
- k 次元潜在ベクトル v の内積で相互作用を表すことで、スパースなデータに対してロバストな学習が可能
- 計算量は $O(kn)$ (n はデータ数)

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

x_i : 入力(説明変数)

w_0 : バイアス

\hat{y} : 出力(目的変数)

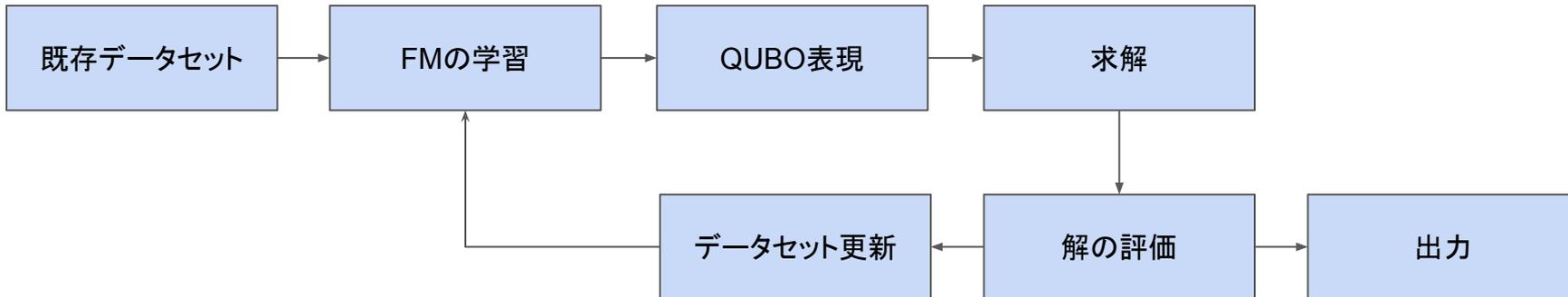
w_i : 変数 i の重み

\mathbf{v}_i : 変数 i 潜在ベクトル

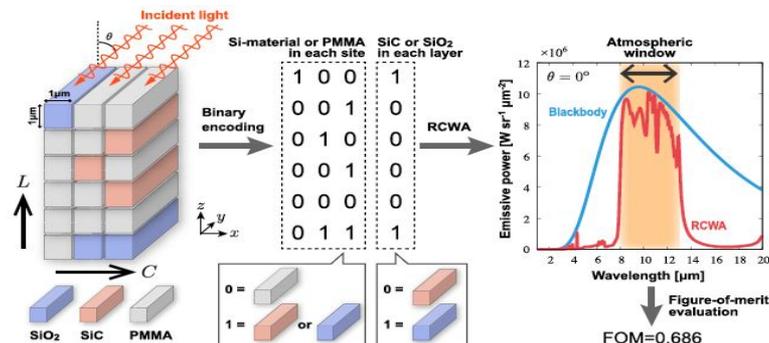
| | Feature vector \mathbf{x} | | | | | | | | | | | | | | | Target y | | | | | |
|-----------|-----------------------------|---|---|-----|-------|----|----|----|-----|--------------------|-----|-----|-----|-----|------|------------------|----|----|----|-----|---|
| $x^{(1)}$ | 1 | 0 | 0 | ... | 1 | 0 | 0 | 0 | ... | 0.3 | 0.3 | 0.3 | 0 | ... | 13 | 0 | 0 | 0 | 0 | ... | 5 |
| $x^{(2)}$ | 1 | 0 | 0 | ... | 0 | 1 | 0 | 0 | ... | 0.3 | 0.3 | 0.3 | 0 | ... | 14 | 1 | 0 | 0 | 0 | ... | 3 |
| $x^{(3)}$ | 1 | 0 | 0 | ... | 0 | 0 | 1 | 0 | ... | 0.3 | 0.3 | 0.3 | 0 | ... | 16 | 0 | 1 | 0 | 0 | ... | 1 |
| $x^{(4)}$ | 0 | 1 | 0 | ... | 0 | 0 | 1 | 0 | ... | 0 | 0 | 0.5 | 0.5 | ... | 5 | 0 | 0 | 0 | 0 | ... | 4 |
| $x^{(5)}$ | 0 | 1 | 0 | ... | 0 | 0 | 0 | 1 | ... | 0 | 0 | 0.5 | 0.5 | ... | 8 | 0 | 0 | 1 | 0 | ... | 5 |
| $x^{(6)}$ | 0 | 0 | 1 | ... | 1 | 0 | 0 | 0 | ... | 0.5 | 0 | 0.5 | 0 | ... | 9 | 0 | 0 | 0 | 0 | ... | 1 |
| $x^{(7)}$ | 0 | 0 | 1 | ... | 0 | 0 | 1 | 0 | ... | 0.5 | 0 | 0.5 | 0 | ... | 12 | 1 | 0 | 0 | 0 | ... | 5 |
| | A | B | C | ... | TI | NH | SW | ST | ... | TI | NH | SW | ST | ... | Time | TI | NH | SW | ST | ... | |
| | User | | | | Movie | | | | | Other Movies rated | | | | | | Last Movie rated | | | | | |

Rendle, S. (2010). Factorization Machines. IEEE International Conference on Data Mining.

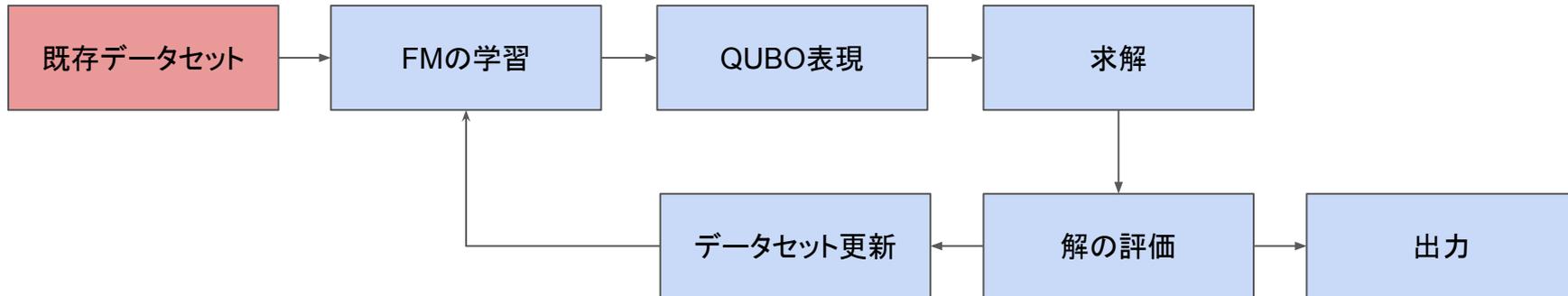
手法の説明



Kitai, K., Guo, J., & Ju, S. (2020). Designing Metamaterials with Quantum Annealing and Factorization Machines. *PHYSICAL REVIEW RESEARCH*, 2(013319).



手法の説明



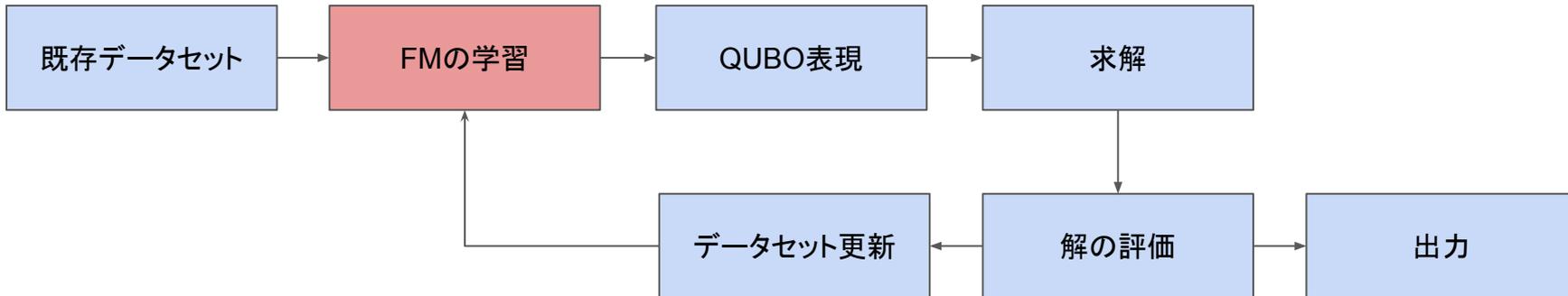
- FMAを実行するには入力(決定変数)と出力(目的変数)が関連付いたデータセットが必要
- 材料開発
 - 入力: 組成・実験条件
 - 出力: 材料の特性
- 深層学習モデルハイパーパラメータ
 - 入力: ハイパーパラメータ
 - 出力: モデルの評価指標

| | act_1 | act_2 | act_3 | hc0_1 | hc0_2 | hc0_3 | norm_1 | norm_2 | score |
|---|-------|-------|-------|-------|-------|-------|--------|--------|----------|
| 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0.611019 |
| 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0.713067 |
| 3 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.691854 |
| 4 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0.693188 |
| 5 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0.701224 |
| 6 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0.634725 |
| 7 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.703049 |

入力

出力

手法の説明

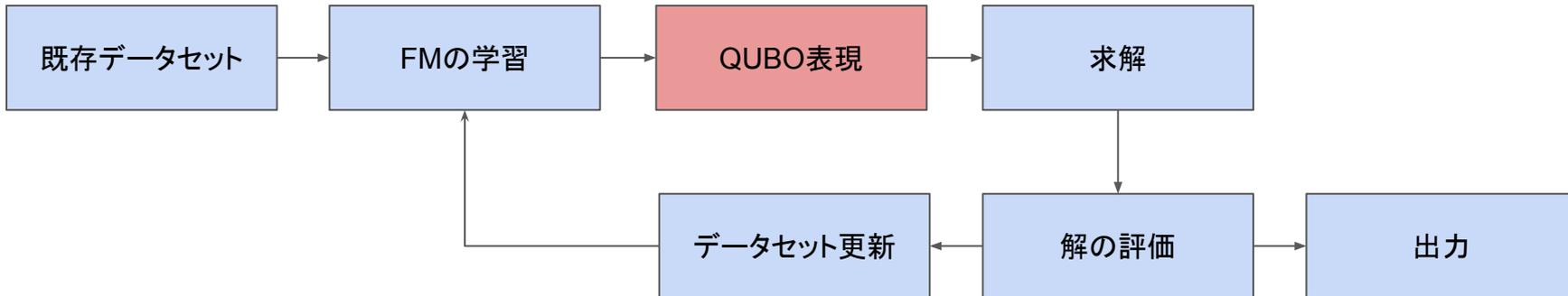


モデル:
$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

- QUBOの一例
- スパースな高次元データに対しても効率的に動作
- 計算量は $O(kN)$

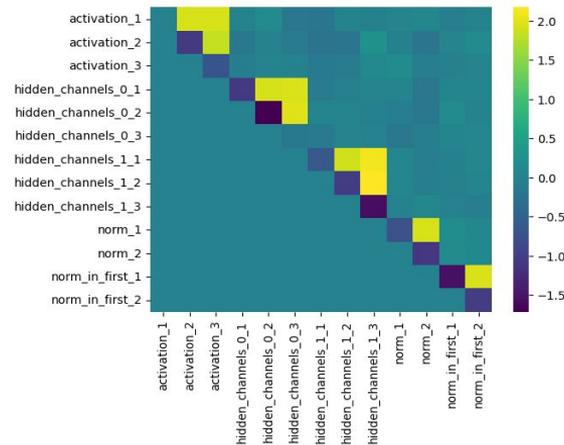
| 変数 | 種類 | 取りうる値 | 次元 |
|----|---------|--------|-------|
| x | 決定変数 | 0 or 1 | n |
| y | 予測値 | 実数 | スカラー |
| w0 | 学習パラメータ | 実数 | スカラー |
| w | 学習パラメータ | 実数 | n |
| v | 学習パラメータ | 実数 | n × k |

手法の説明



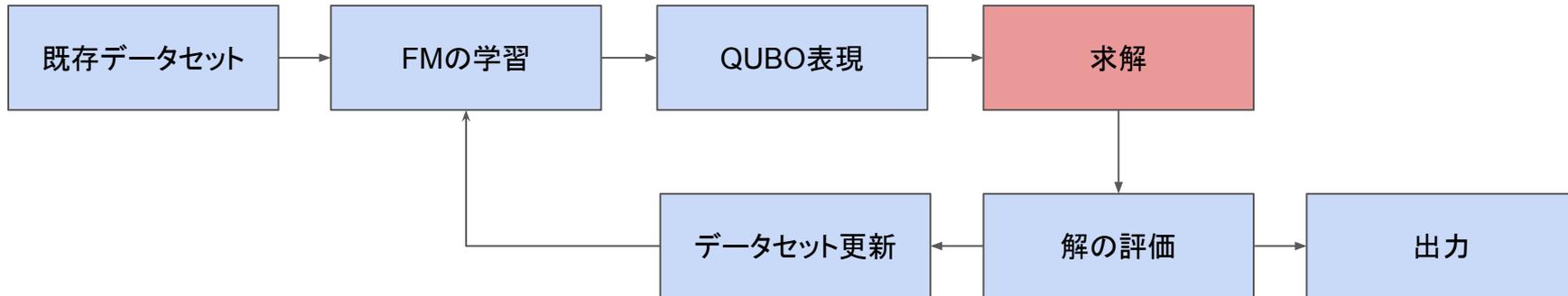
$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^N \sum_{j=i}^N Q_{ij} x_i x_j$$



- FMの学習パラメータからQUBO行列を取得

手法の説明



FMで近似した目的関数を最小化するような 0 or 1の組合せをアニーリングマシンで求める。

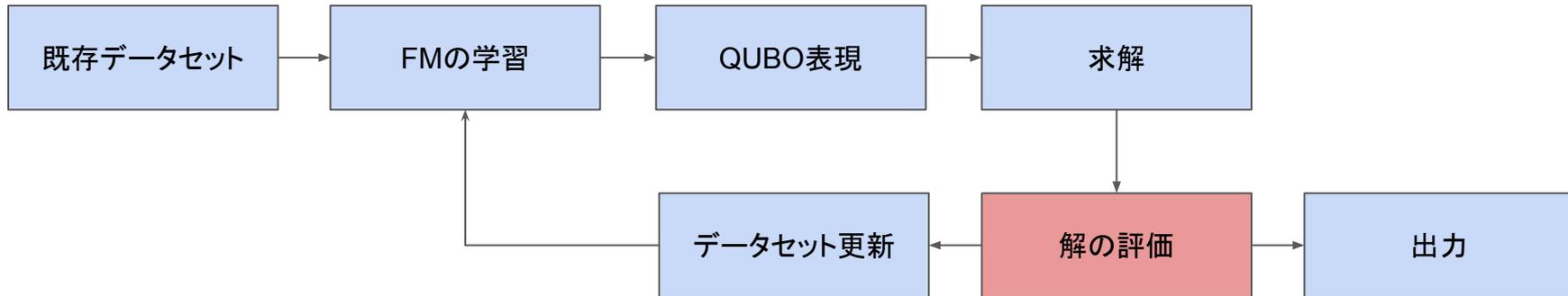
13変数の問題の解の一例

```
[0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0]
```

アニーリングマシン

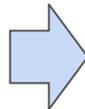
- 量子アニーリング: D-Wave
- Simulated Annealing(OpenJij); CPU
- Digital Annealer (富士通); ASIC, GPU
- SQBM+ (東芝); GPU, FPGA
- SX-Aurora Annealer (NEC); SX-Aurora

手法の説明



- 0 or 1のバイナリの組合せを人が分かる表現へ変換
- 新しい入力に対する出力を評価

```
[0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0]
```

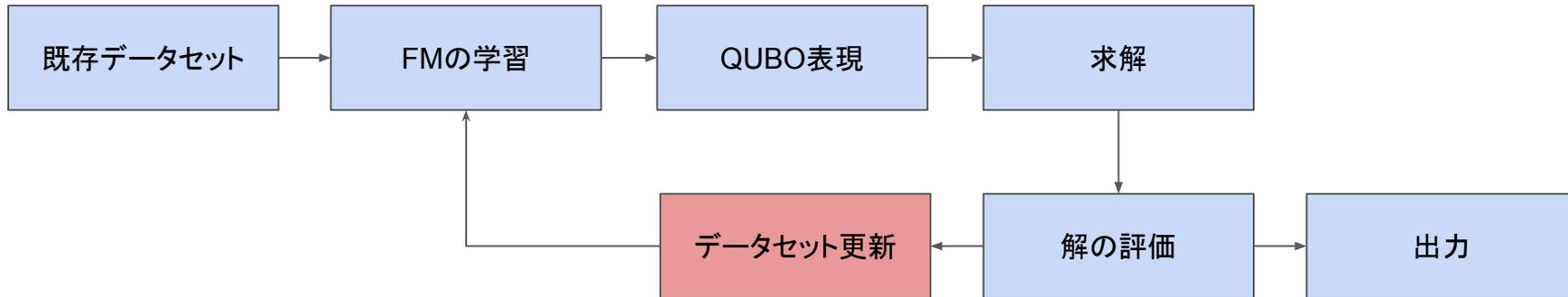


```
["GERU", 24, 48, "LayerNorm", False]
```

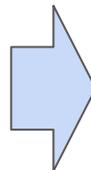
```
Micro Accuracy: 0.7035
```

取得した新規ハイパーパラメータでモデルを学習し性能を評価

手法の説明



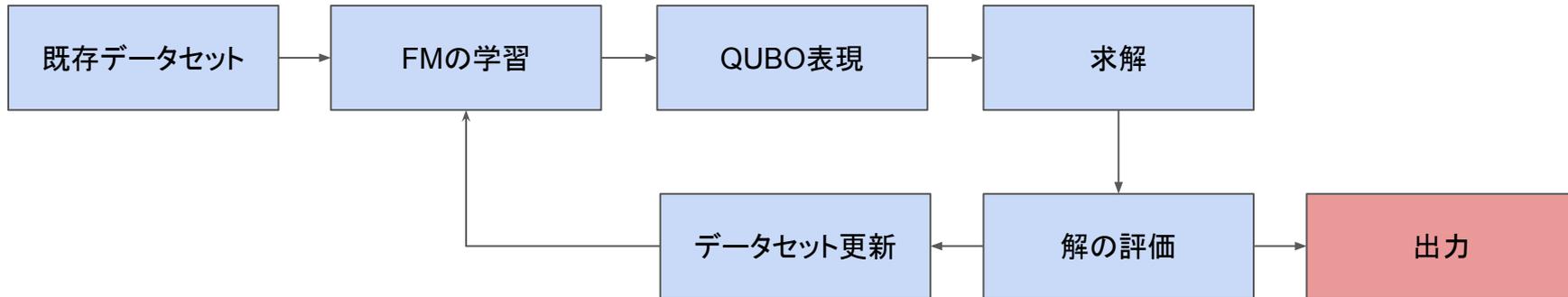
| | act_1 | act_2 | act_3 | hc0_1 | hc0_2 | hc0_3 | norm_1 | norm_2 | score |
|---|-------|-------|-------|-------|-------|-------|--------|--------|----------|
| 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0.611019 |
| 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0.713067 |
| 3 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.691854 |
| 4 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0.693188 |
| 5 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0.701224 |
| 6 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0.634725 |
| 7 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.703049 |



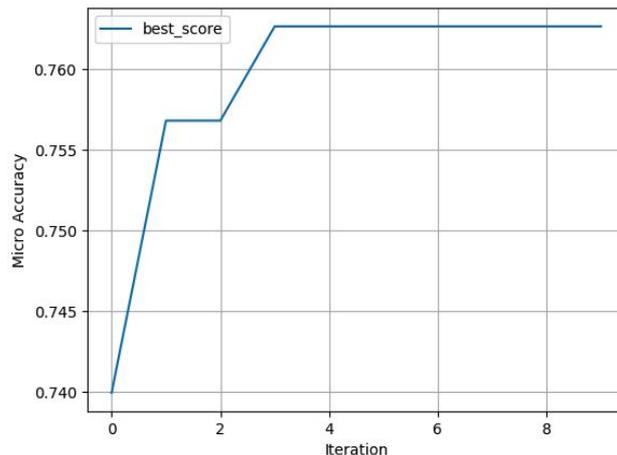
新規データを挿入

| | act_1 | act_2 | act_3 | hc0_1 | hc0_2 | hc0_3 | norm_1 | norm_2 | score |
|---|-------|-------|-------|-------|-------|-------|--------|--------|----------|
| 8 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0.697535 |
| 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0.611019 |
| 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0.713067 |
| 3 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.691854 |
| 4 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0.693188 |
| 5 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0.701224 |
| 6 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0.634725 |
| 7 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.703049 |

手法の説明



- 終了条件まで繰り返し、目的関数値が最小(最大)となる解を出力
- 終了条件
 - 指定の反復回数
 - 目的関数値の改善が見られなくなる迄



ハイパーパラメータ探索

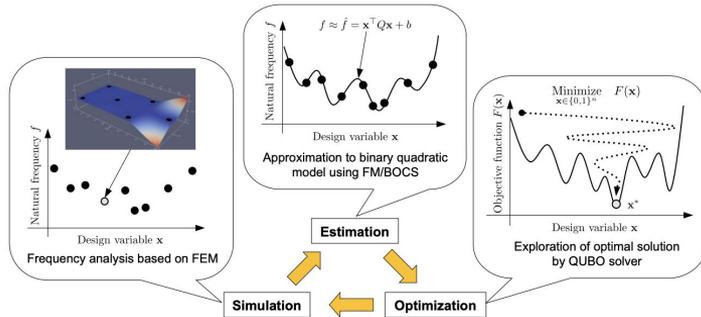
- 機械学習モデルの性能を最大化するような最適なハイパーパラメータを探索する。
- 目的関数はハイパーパラメータを受け取り、学習済みモデルの性能を出力する関数
- ブラックボックス最適化の代表的な問題設定で、ベイズ最適化がよく使われる。

表: ハイパーパラメータ探索の代表的な手法と課題

| 手法名 | 概要 | 課題 |
|---------|---|---|
| グリッドサーチ | <ul style="list-style-type: none">● 予め定義された複数のハイパーパラメータの値の組み合わせを全て試す方法 | <ul style="list-style-type: none">● 計算コストが非常に高い。ハイパーパラメータが多い場合に現実的に計算不可能となる。 |
| ランダムサーチ | <ul style="list-style-type: none">● ハイパーパラメータの値をランダムに選択して試す方法 | <ul style="list-style-type: none">● 効率的な探索が難しく、最適解にたどり着く保証がない。 |
| ベイズ最適化 | <ul style="list-style-type: none">● サロゲートモデルで目的関数を近似し、その目的関数から構築される獲得関数を最適化することで、ハイパーパラメータ空間を探索する方法● グリッドサーチ、ランダムサーチに比べ効率的 | <ul style="list-style-type: none">● サロゲートモデルにガウス過程を用いる方法では離散変数の取り扱いが難しい。● TPEは推定のばらつきが大きい● ハイパーパラメータ空間が高次元になると計算コストが高くなる。 |

本研究の目的

- FMAを機械学習モデルのハイパーパラメータ探索に適用することで、ハイパーパラメータ探索の最適化手法としての可能性を検証する。
- 検証は以下の観点で実施する。
 1. 他手法との比較
 2. FMAの性能調査(計算効率、スケーラビリティ)



Tadayoshi Matsumori, Masato Taki & Tadashi Kadowaki, "Application of QUBO solver using black-box optimization to structural design for resonance avoidance", Scientific Reports 12, 12143 (2022) より引用

| <i>Hyperparameters</i> | <i>Investigated values</i> |
|--------------------------|--|
| N_FEATURES (LightGBM) | [96, 72, 48, 24, 12] |
| N_HIDDEN_LAYER (LSTM) | [12, 24, 36, 48, 64, 78, 96, 108, 120] |
| N_LAYERS (TRANSFORMER) | [2, 4, 6, 8, 10, 12, 14] |
| CONV_ID_HIDDEN_CHANNEL | [[12, 12], [24, 12], [12, 24], [24, 24], [48, 12], [48, 24], [12, 48], [24, 48], [48, 48]] |
| ACTIVATION | [ReLU, Tanh, GELU] |
| NORM | [BatchNorm, LayerNorm] |
| NORM_IN_FIRST_CONV_BLOCK | [True, False] |

M.Shirotori, K.Kosuke, T.Hondo, K.Kim, K.Tokuyama, "Improved generalized performance of Hemodynamics Scenarios prediction with digital biomarkers by Conv1D approach", IEEE SMC 2023より引用

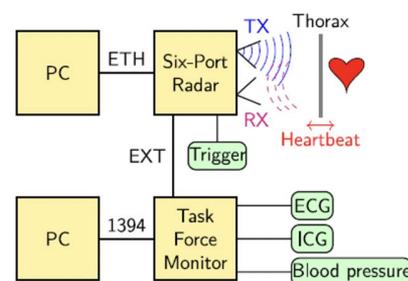
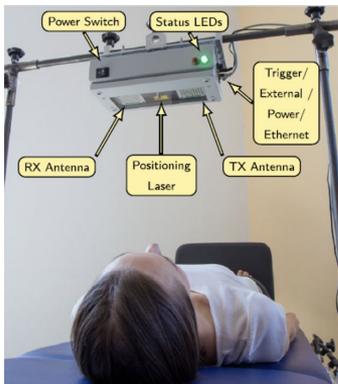
問題設定

血行動態

- 血液の流れやその流れが生体組織に与える力の研究
- 血圧、血流量、血管の抵抗、心拍出量など

重要性

- 疾病の診断: 血行動態の変化は多くの心血管疾患の早期指標となり得る。例えば、動脈硬化は血管の弾力性を低下させ、血行動態に影響を及ぼす。
- 治療の指針: 血行動態の評価は、心不全や高血圧といった状態の管理において治療計画を立てる上で重要となる。例えば、薬物による血圧の調整や、血流を改善する手術の検討材料となる。
- 予防医学: 血行動態の正常範囲を知ることで、予防措置としてのライフスタイルの調整や早期介入が可能となり、循環器疾患のリスクを下げることができる。



S. Schellenberger, K. Shi, T. Steigleder, A. Malessa, F. Michler, L. Hameyer, N. Neumann, F. Lurz, R. Weigel, C. Ostgathe, and A. Koelpin, "A dataset of clinically recorded radar vital signs with synchronised reference sensor signals," Sci. Data Vol. 7, p.291, 2020.

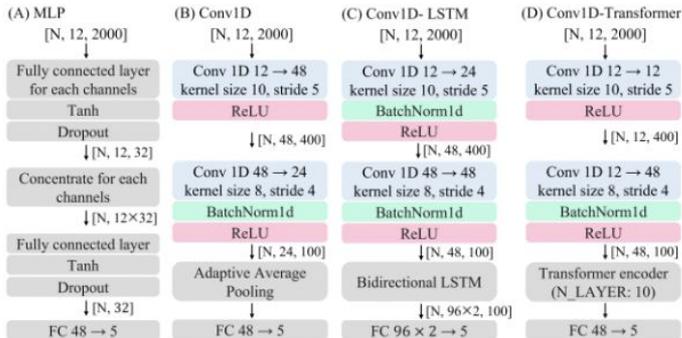
問題設定

デジタルバイオマーカー

- デジタルバイオマーカーはウェアラブルデバイスやスマートフォンのアプリケーションを使用して得られるデータ(心拍数、睡眠パターン、血糖レベルなど)を用いて、健康状態を客観的に評価する指標を指す。
- 連続的でリアルタイムのデータ提供が可能であり、より個別化された健康管理や予防医学への貢献など医療の質を向上させる新しい方法として注目されている。

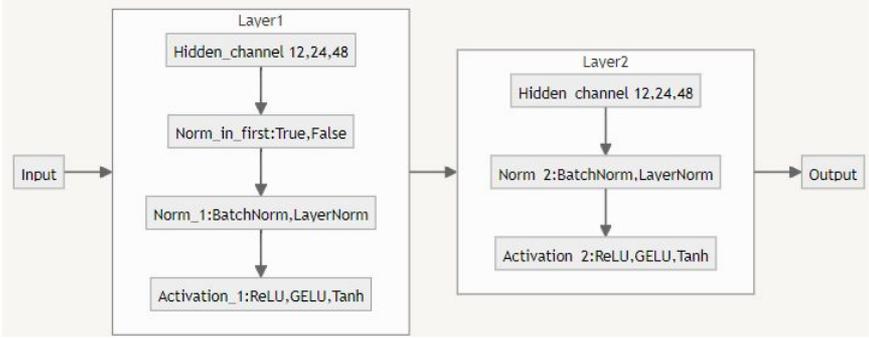
デジタルバイオマーカーを活用した血行動態予測の先行研究

- 接触型および非接触型センサーから得られた12種類の生体信号を用いて、血圧変化や心拍数の変化など5種類の血行動態シナリオを分類する。
- 分類モデルには1次元畳み込みニューラルネットワークを使用している。



| | <i>Micro-accuracy</i> | <i>Macro-recall</i> | <i>Macro-precision</i> | <i>Macro-F1</i> | <i>Macro-AUC</i> |
|--------------------|-----------------------|---------------------|------------------------|-----------------|------------------|
| MLP | 0.436 | 0.214 | 0.217 | 0.213 | 0.535 |
| LightGBM | 0.709 | 0.617 | 0.700 | 0.631 | 0.902 |
| Conv1D | 0.745 | 0.788 | 0.644 | 0.671 | 0.936 |
| Conv1D-BiLSTM | 0.771 | 0.763 | 0.677 | 0.695 | 0.930 |
| Conv1D-Transformer | 0.747 | 0.770 | 0.632 | 0.653 | 0.935 |

問題設定

| | |
|--------------------------|--|
| <p>先行研究</p> | <p>血行動態シナリオ(Resting,the Valsalva manoeuvre,apena(息止め), the tilt up table test,the tilt down table test)の分類モデルとして幾つかの深層学習モデルを作り、その精度を比較した研究^[1]を公表</p> |
| <p>深層学習モデルに用いるデータセット</p> | <p>30人の健康な被験者から収集した数十種類の時系列センサーデータ。上記の5つの血行動態シナリオを実施中に各シナリオで数十分間データを取得^[2]。</p> |
| <p>目的関数値</p> | <p>Conv1Dベースのモデル^[1]について、その分類精度であるMicro accuracyを最大化</p> |
| <p>決定変数</p> | <p>Hidden_channel; [12, 12], [24, 12], [12, 24],[24, 24], [48, 12], [48, 24],[12, 48], [24, 48], [48, 48] Norm_in_first; True, False Norm;BatchNorm, LayerNorm Activation;ReLU, Tanh, GELU ※合計の組合せ総数は108(=9*2*2*3)通り</p>  |

[1]Shirotori, M., Kosuke, K., Hondo, et. al, (2022). Improved generalized performance of Hemodynamics Scenarios prediction with digital biomarkers by Conv1D approach.

[2]Schellenberger, S., Shi, K., Steigleder, T., Malessa, et. al, (2020). A dataset of clinically recorded radar vital signs with synchronised reference sensor signals. Scientific Data, 7, 5.

検証用データセット作成

$$\hat{H}(\sigma) = A\hat{E}_1 + B\hat{E}_2 + \dots$$

$$\hat{E}_1 = \sum_{i=1}^N \left(\sum_{j=1}^N \hat{\sigma}_{i,j} \right)$$

$$\hat{E}_2 = \sum_{i=1}^N \left(\sum_{j=1}^N \hat{\sigma}_{i,j} \right)$$

$$\hat{E}_3 = \sum_{i,j,k=1}^N J_{i,j,k}$$

グリッドサーチによるデータセット作成

| | |
|-------------------------|---|
| 目的 | FMAの性能を検証するためのデータセット作成 |
| 概要 | Conv1Dモデルを実装し、108、300、1200通りのハイパーパラメータパターンについてそれぞれグリッドサーチを実施 変数の数はそれぞれ13、17、27 |
| Conv1Dモデルの学習データ | 20人の被験者データ※先行研究[1]では30人分のデータを使用 |
| Conv1Dモデルのハイパーパラメータ（固定） | 先行研究のConv1Dモデルに合わせて実装。主要なパラメータは以下。 <ul style="list-style-type: none">● kernel size = 3● learning rate = 0.001● batch size = 32● epoch = 50 |
| クラス分類の評価値 | Micro accuracy (= TP / (TP+FP+FN)) |

[1]Shirotori, M., Kosuke, K., Hondo, et. al, (2022). Improved generalized performance of Hemodynamics Scenarios prediction with digital biomarkers by Conv1D approach.

ランダムサーチによるデータセット作成

| | |
|--------------------------|---|
| 目的 | 最適値が不明な場合のFMAの性能を検証するためのデータセット作成 |
| 概要 | Conv1Dモデルを拡張したモデルを実装し、ランダムサーチを100回実施 ハイパーパラメータパターンは 10^{10} 変数の数は84 |
| 学習データ | 20人の被験者データ |
| Conv1Dモデルのハイパーパラメータ (固定) | 先行研究のConv1Dモデルに合わせて実装。主要なパラメータは以下。 <ul style="list-style-type: none">● learning rate = 0.001● batch size = 32● epoch = 50 |
| クラス分類の評価値 | Micro accuracy (= TP / (TP+FP+FN)) |

グリッドサーチによるデータセット作成

パラメータパターン数:108
データセット名:DS108
最適値:0.7655

| ハイパーパラメータ名 | パラメータ候補 |
|-------------------|------------------------------|
| activation | ["ReLU", "Tanh", "GELU"] |
| hidden_channels_0 | [12, 24, 48] |
| hidden_channels_1 | [12, 24, 48] |
| norm | ["BatchNorm1d", "LayerNorm"] |
| norm_in_first | [True, False] |

パラメータパターン数:300
データセット名:DS300
最適値:0.7655

| ハイパーパラメータ名 | パラメータ候補 |
|-------------------|------------------------------|
| activation | ["ReLU", "Tanh", "GELU"] |
| hidden_channels_0 | [12, 24, 48, 64, 128] |
| hidden_channels_1 | [12, 24, 48, 64, 128] |
| norm | ["BatchNorm1d", "LayerNorm"] |
| norm_in_first | [True, False] |

パラメータパターン数:1200
データセット名:DS1200
最適値:0.7756

| ハイパーパラメータ名 | パラメータ候補 |
|-------------------|---|
| activation | ["ReLU", "Tanh", "GELU"] |
| hidden_channels_0 | [8, 12, 24, 32, 48, 64, 80, 96, 112, 128] |
| hidden_channels_1 | [8, 12, 24, 32, 48, 64, 80, 96, 112, 128] |
| norm | ["BatchNorm1d", "LayerNorm"] |
| norm_in_first | [True, False] |

ランダムサーチによるデータセット作成

パラメータパターン数 :380億

データセット名 :DS38bil

最良値 :0.9777(ランダムサーチを100回繰り返したときに見つかった最良の値)

| ハイパーパラメータ名 | パラメータ候補 |
|-------------------|--|
| activation_1 | ["ReLU", "Tanh", "GELU", "LeakyReLU"] |
| activation_2 | ["ReLU", "Tanh", "GELU", "LeakyReLU"] |
| activation_3 | ["ReLU", "Tanh", "GELU", "LeakyReLU"] |
| dropout_1 | [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0] |
| dropout_2 | [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0] |
| dropout_3 | [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0] |
| hidden_channels_0 | [12, 24, 32, 48, 64, 96, 128] |
| hidden_channels_1 | [12, 24, 32, 48, 64, 96, 128] |
| hidden_channels_2 | [12, 24, 32, 48, 64, 96, 128] |
| norm | ["BatchNorm1d", "LayerNorm", "Identity"] |
| norm | ["BatchNorm1d", "LayerNorm", "Identity"] |
| norm | ["BatchNorm1d", "LayerNorm", "Identity"] |
| kernel_size_1 | [1, 3, 5, 10] |
| kernel_size_2 | [1, 3, 5, 10] |
| kernel_size_3 | [1, 3, 5, 10] |

FMAの性能検証

$$\hat{H}(\sigma) = A\hat{E}_1 + B\hat{E}_2 + \dots$$

$$\hat{E}_1 = \sum_{i=1}^N \left(\sum_{j=1}^N \hat{\sigma}_{i,j} \right)$$

$$\hat{E}_2 = \sum_{i=1}^N \left(\sum_{j=1}^N \hat{\sigma}_{i,j} \right)$$

$$\hat{E}_3 = \sum_{i,j,k=1}^N J_{i,j,k}$$

FMAの実験条件

| | |
|---------------|--|
| 実験概要 | <ol style="list-style-type: none">DS108、DS300、DS1200を用いて、最適解に収束するまでのFMAの探索回数を調査より大きいハイパーパラメータ空間(DS300,DS1200)において、1と同様の実験を実施最適解に到達するまでの探索回数の初期学習データ数依存性 |
| 学習データの初期サンプル数 | <ul style="list-style-type: none">実験1,2については初期学習データ数= 20実験3については[2, 4, 8, 10, 15, 20, 30] |
| FMの学習条件 | 以下のパラメータを各反復処理で最適化 <ul style="list-style-type: none">FMモデルの学習における反復回数(n_iter)FMモデルの表現力を表すパラメータ(rank)FMモデルパラメータの初期化に使用される標準偏差(nit_stddev)1次正則化係数(l2_reg_w)2次正則化係数(l2_reg_V) |
| 使用ソルバー | SA(OpenJij)、QA(Leap Hybrid CQM)、Optuna |
| FMAの終了条件 | 繰り返し回数 = 20 ~ 100 |
| 備考 | DSxxx: xxx通りのハイパーパラメータとMicro Accuracyからなるデータセットの名前 |

FMAの性能検証

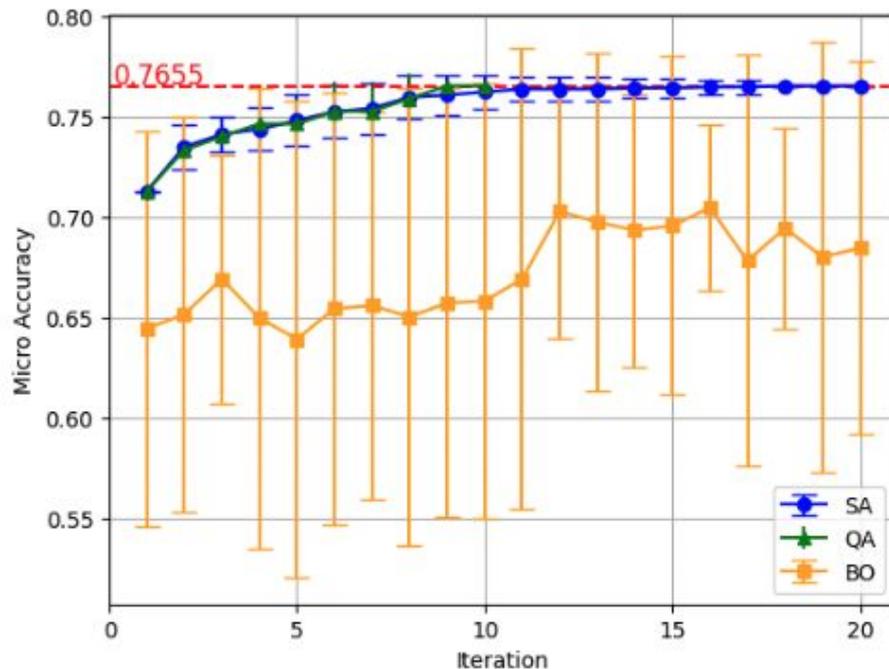
実験条件

- 検証データ
 - DS108
 - 初期サンプル数：20
- 手法
 - FM+SA(OpenJij)
 - FM+QA (Leap Hybrid CQM)
 - ベイズ最適化 (Optuna)
- 試行回数
 - SA：50回
 - QA：10回
 - BO：50回

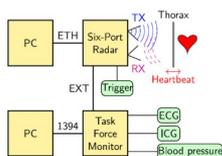
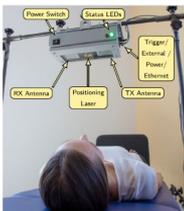
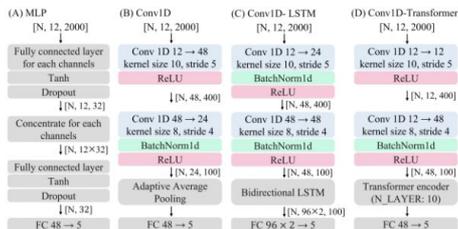
結果

- SA, QA共にでは約10回の探索回数で最適解へ収束するが、QAの方が少しだけ早い。
- 初期サンプル数を考慮すると、グリッドサーチと比較して約70%少ない探索回数で最適値を発見出来る。
- ベイズ最適化より収束性能は良さそう。
- 一回あたりの探索時間は数秒程度

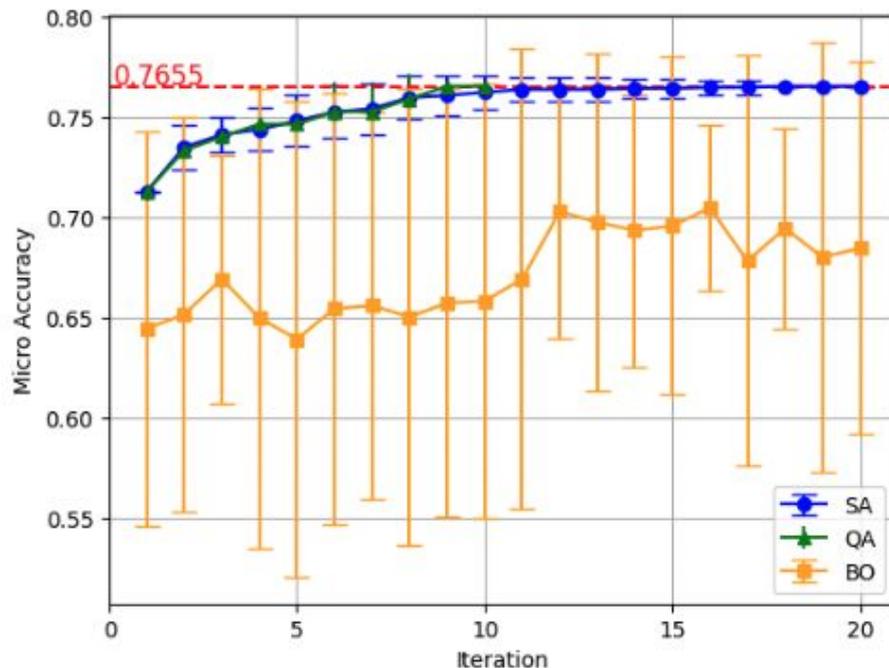
FMAにおける目的関数の推移（検証データ：DS108）



血行動態予測のための NNアーキテクチャの最適化： 中外製薬様

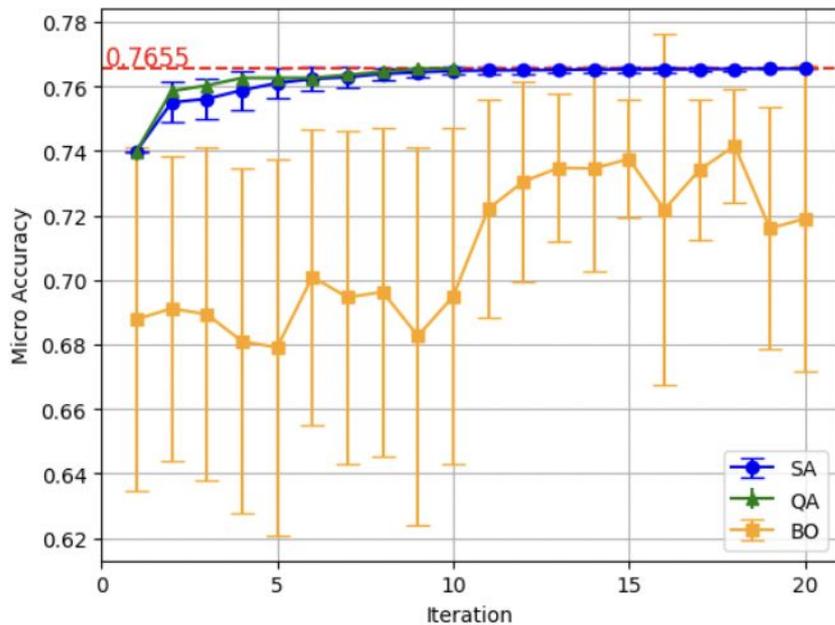


S. Schelemburger, K. Shi, T. Stegleder, A. Malessa, F. Michter, L. Hameyer, N. Neumann, F. Lurz, R. Weigel, C. Ostgathe, and A. Koelpin, "A dataset of clinically recorded radar vital signs with synchronised reference sensor signals," Sci. Data Vol. 7, p.201, 2020.

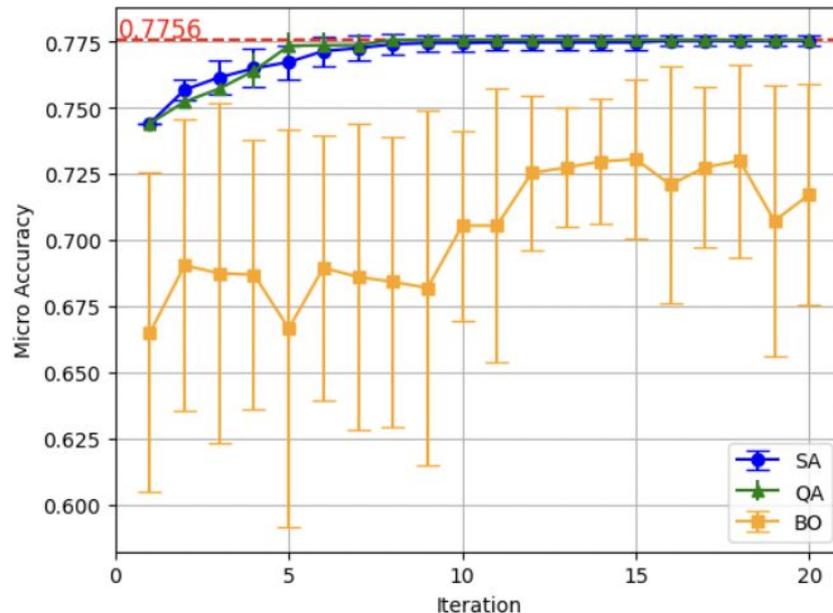


FMAの性能検証

FMAにおける目的関数の推移 (検証データ : DS300)



FMAにおける目的関数の推移 (検証データ : DS1200)



FMAの性能検証

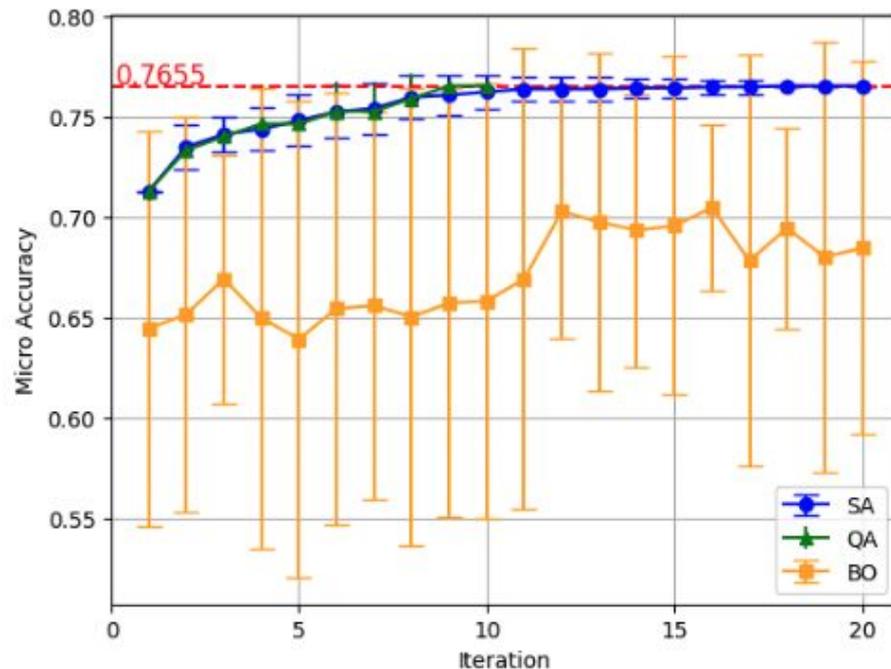
実験条件

- 検証データ
 - DS108
 - 初期サンプル数：20
- 手法
 - FM+SA(OpenJij)
 - FM+QA (Leap Hybrid CQM)
 - ベイズ最適化 (Optuna)
- 試行回数
 - SA：50回
 - QA：10回
 - BO：50回

結果

- SA, QA共にでは約10回の探索回数で最適解へ収束するが、QAの方が少しだけ早い。
- 初期サンプル数を考慮すると、グリッドサーチと比較して約70%少ない探索回数で最適値を発見出来る。
- ベイズ最適化より収束性能は良さそう。
- 一回あたりの探索時間は数秒程度

FMAにおける目的関数の推移（検証データ：DS108）



FMA for ds300,ds1200の実験結果

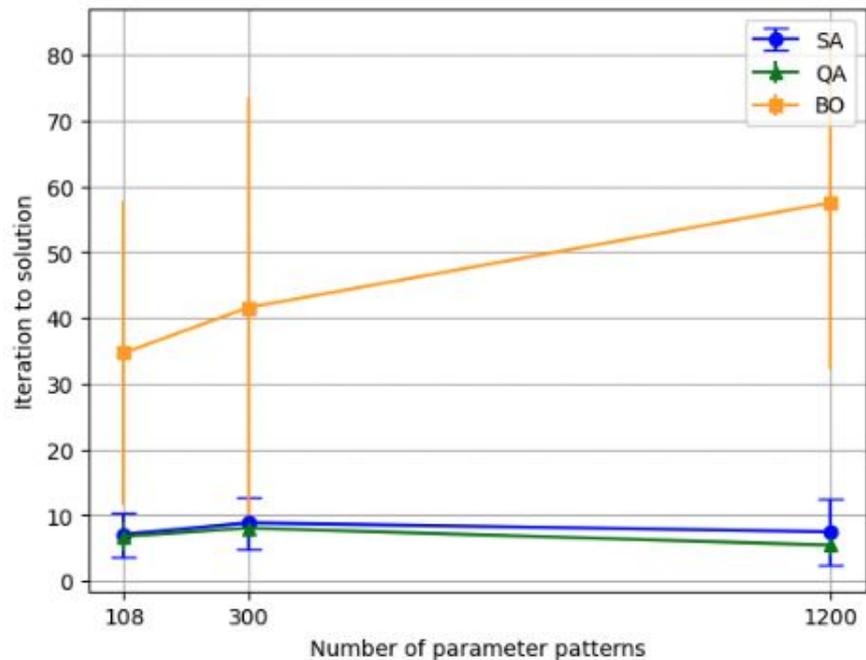
実験条件

- 検証データ
 - DS108, DS300, DS1200
 - 初期サンプル数：20
- 手法
 - FM+SA(OpenJij)
 - FM+QA (Leap Hybrid CQM)
 - ベイズ最適化 (Optuna)
- 試行回数
 - SA：10回
 - QA：5回
 - BO：10回

結果

- QAとSAはパラメータ数が増えても探索回数に平均の探索回数はほとんど変わらない。
- ベイズ最適化はパラメータ数が増えると探索回数も増加

最適解に到達するまでの探索回数



探索回数の初期サンプル数依存性

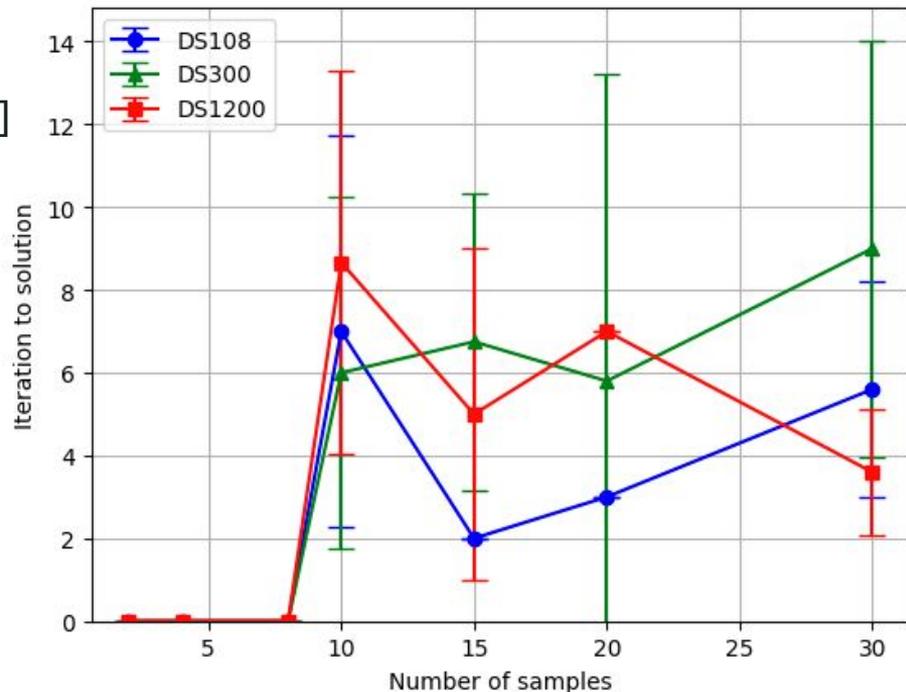
実験条件

- 学習データ
 - DS108, DS300, DS1200
 - 初期サンプル数：[2, 4, 8, 10, 15, 20, 30]
- ソルバー
 - SA
- FMAの反復回数
 - 30回
- 試行回数
 - 10回

結果

- 初期サンプル数8まではFMの学習データが足りず最適化が進まないが、初期サンプル数10のとき初めて最適解へ収束する。

図4:解探索における初期学習データの影響



$$\hat{H}(\sigma) = A\hat{E}_1 + B\hat{E}_2 + \dots$$

$$\hat{E}_1 = \sum_{i=1}^N \left(\sum_{j=1}^N \hat{\sigma}_{i,j} \right)$$

$$\hat{E}_2 = \sum_{i=1}^N \left(\sum_{j=1}^N \hat{\sigma}_{i,j} \right)$$

$$\hat{E}_3 = \sum_{i,j,k=1}^N J_{i,j,k}$$

ハイパーパラメータの大規模化の実験

FMSA for 拡張Conv1Dモデルの実験条件

| | |
|---------------|--|
| 実験概要 | <ol style="list-style-type: none">1. 拡張Conv1Dモデルに対してランダムサーチ(100回)とFMSAを実施及び、最適解を得るまでの探索回数を比較2. 最適解に到達するまでの探索回数の初期学習データ数依存性3. ベイズ最適化による探索を実施 |
| 学習データの初期サンプル数 | <ul style="list-style-type: none">● 実験1については初期学習データ数= 20● 実験2については初期学習データ数= [2, 4, 8, 20, 35, 50] |
| FMの学習条件 | 以下のパラメータを各反復処理で最適化 <ul style="list-style-type: none">● FMモデルの学習における反復回数(n_iter)● FMモデルの表現力を表すパラメータ(rank)● FMモデルパラメータの初期化に使用される標準偏差(nit_stddev)● 1次正則化係数(l2_reg_w)● 2次正則化係数(l2_reg_V) |
| 使用ソルバー | SA(OpenJij)、QA(Leap Hybrid CQM)、Optuna |
| FMAの終了条件 | 繰り返し回数 = 50 |
| 備考 | FMに追加する学習データは都度都度Conv1Dモデルで計算 ランダムサーチを100回行った中の最適値は0.9777 (詳細は付録参照) |

FMSA for 拡張Conv1Dモデル

- 2層→3層へ拡張
- activationの種類 ["ReLU", "Tanh", "GELU", "LeakyReLU"]
- dropoutを追加 [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
- ネットワークサイズ [12, 24, 32, 48, 64, 96, 128]
- kernel_size [1, 3, 5, 10]
- 正規化手法 ["BatchNorm1d", "LayerNorm", "Identity"] ※赤字はDS108と異なる部分

総組合せ数は約380億
FMの決定変数の数は84

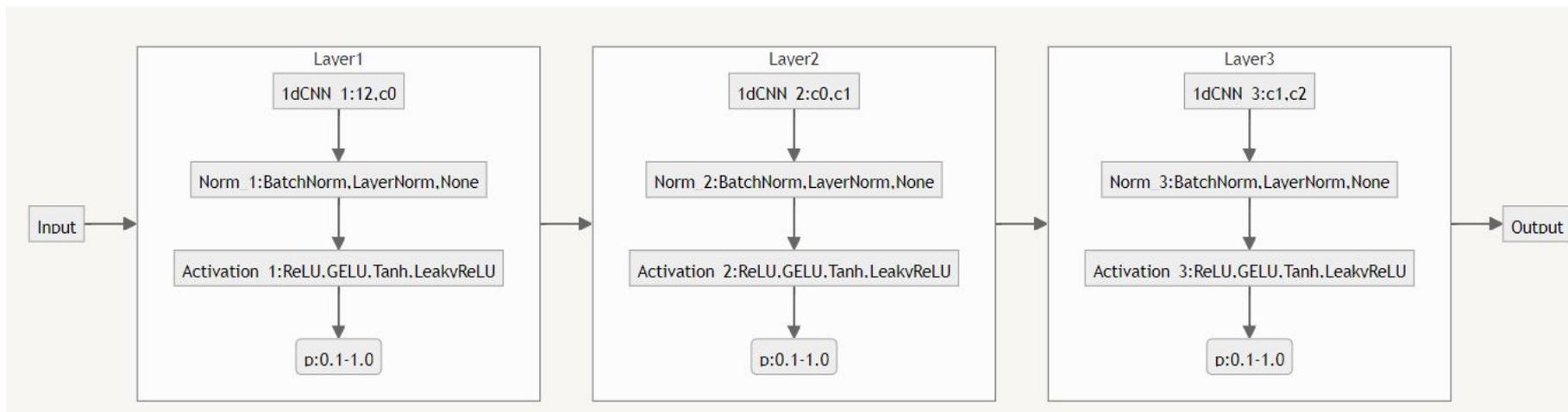


図5:拡張Conv1Dモデルのネットワークの詳細

FMSA for 拡張Conv1Dモデルの実験結果

- FMSAだと22回目※の探索でランダムサーチで見つかった最大値(0.9777)を超える値(0.9799)を発見
- 一回あたりの探索時間はベイズ最適化: 約60秒、FMA: 約10秒

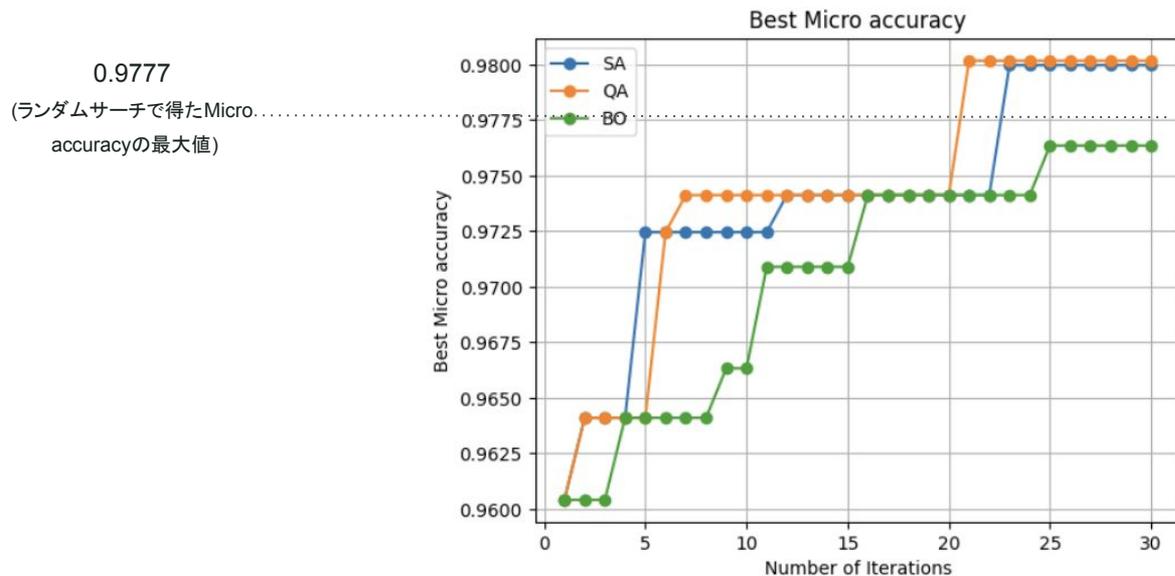


図6:FMSAで探索したmicro accuracyの推移

※初期学習データにを使った20回を考慮すると、正確には42回のCNNモデル学習・評価で該当の値を発見

FMSA for 拡張1dCNNモデルの実験結果

FMへの初期学習データ量が少ない※方が、少ない実質合計探索回数($n_samples$ を含む回数)で早く目的の値(0.9777)を見つけることが出来る傾向が見られる。

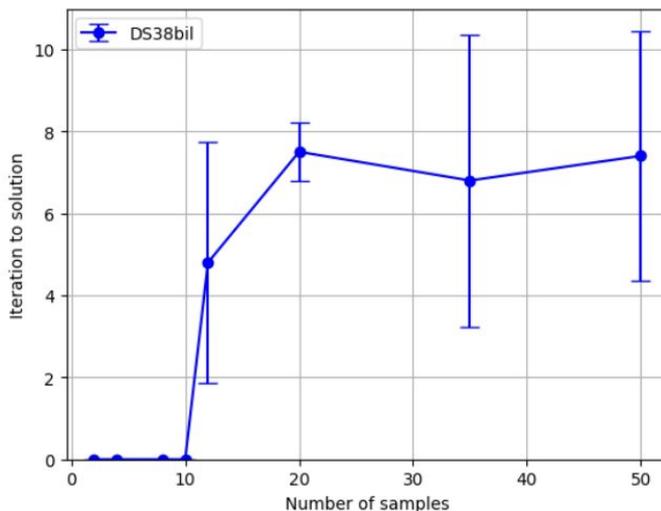


図7:FMSAの初期学習データ($n_samples$)と、ランダムサーチの最適値を見つけるまでのFMSAの探索回数の関係

※ $n_samples$ は最低12以上ないとFMモデルは目的の値を探索出来ないのもそれ以上の値は必要。

まとめと今後の展望

本研究では、FMA (Factorization Machine with Annealing) を機械学習モデルのハイパーパラメータ探索に適用し、その効果をベイズ最適化と比較した。主な発見は以下の通り。

1. **スケーラビリティ**: FMAはハイパーパラメータパターンが増加しても、ロバストな性能を発揮する。
2. **計算時間**: FMAはベイズ最適化と比較して、1回あたりの探索時間が短いことが確認された(ベイズ最適化: 約60秒、FMA: 約10秒)。SAとQAであまり性能差はなくSAでも十分

今後の展望

- 初期サンプル数依存性の調査
- 他の機械学習モデルでの性能検証を行い、FMAの汎用性をさらに確認する。
- NAS (Neural Architecture Search) など、より広範囲の探索を必要とする問題へ適用する。