2021/8/5 @土木学会

機械学習モデルによる地形情報からの工学的基盤深度の推定モデル構築

公益財団法人鉄道総合技術研究所　田中 浩平

発表内容

A. 受賞論文について

B. 今後の地震工学における機械学習の活用について

A.1.1 背景

- 自然災害分野におけるハザードマップ
- 地形的要因と災害発生原因に関する専門的な知見
  - 微地形区分（若松・松岡（2008）他）
  - 液状化危険度（松岡ら（2011）、中埜（2020））
  - 斜面崩壊（周ら（2008）、木下（2019））
- 日本全国を対象とした場合に整備に要する多大なコスト、専門家によるルール化の限界、高度な専門化・暗黙知化が将来にわたっての持続的・分野横断的な展開の障害となりうる

A.1.2 目的

- 防災分野における機械学習的アプローチの有効性
- 親和性
  - 自然現象にはルール化できない複雑な背後構造
  - 1次データが十分に蓄積され、公開されているものも多い
    （今後も増え続ける）
  - 継続的な更新が必要であり、コストが掛かる
- 地形情報に基づく工学的基盤深度の推定モデルを構築

A.2 推定モデルの構築

入力値：地形情報から作成する各種特徴量

出力値：工学的基盤面の深度(m)

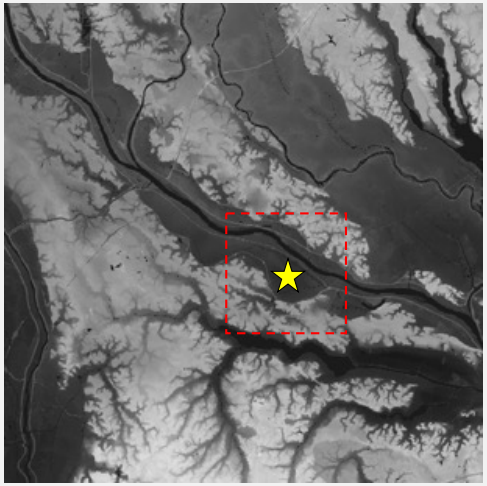
※工学的基盤面は，Vs=400m/sもしくはN値50が連続する地層の上面

構築手順

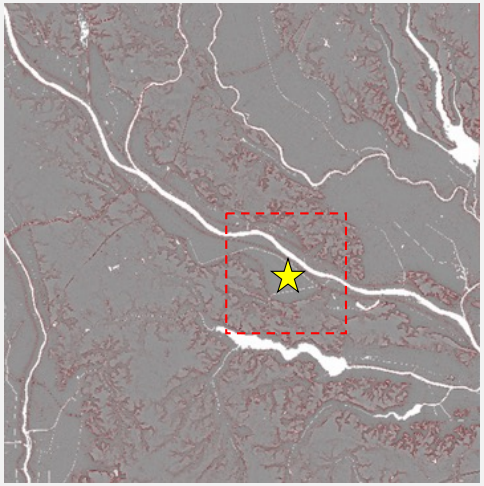
① タスクの種類の設定（分類、回帰）
② モデルの評価指標の設定（推定深度の誤差RMS）
③ データの収集・整理（A.2.1）
④ 特徴量の作成（A.2.2）
⑤ モデルの作成（画像、表形式、A.2.3）
⑥ モデルの評価（A.2.4）

A.2.1 データの収集・整理

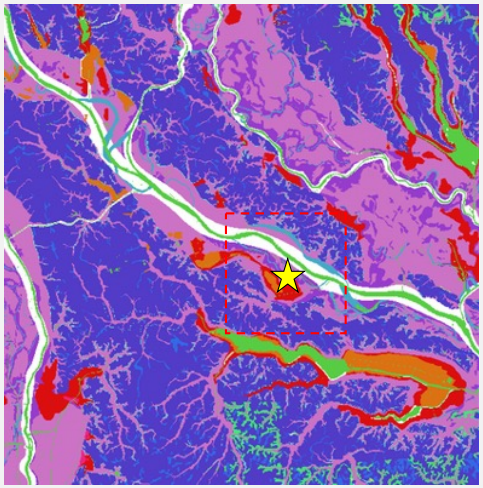
地形情報：任意の解像度の点・メッシュごとに値を有するデータ



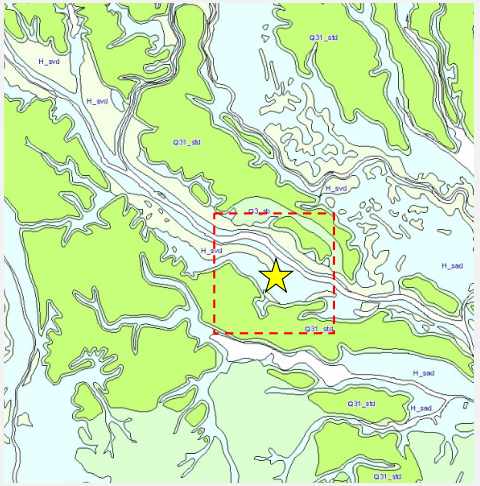
(a) DEM



(b) 赤色立体地図



(c) 地形分類



(d)シームレス地質図

(a), (b), (c) 国土地理院、(d) 産総研

※評価地点☆だけでなく、その周辺 の地形情報も活用する

A.2.1 データの収集・整理

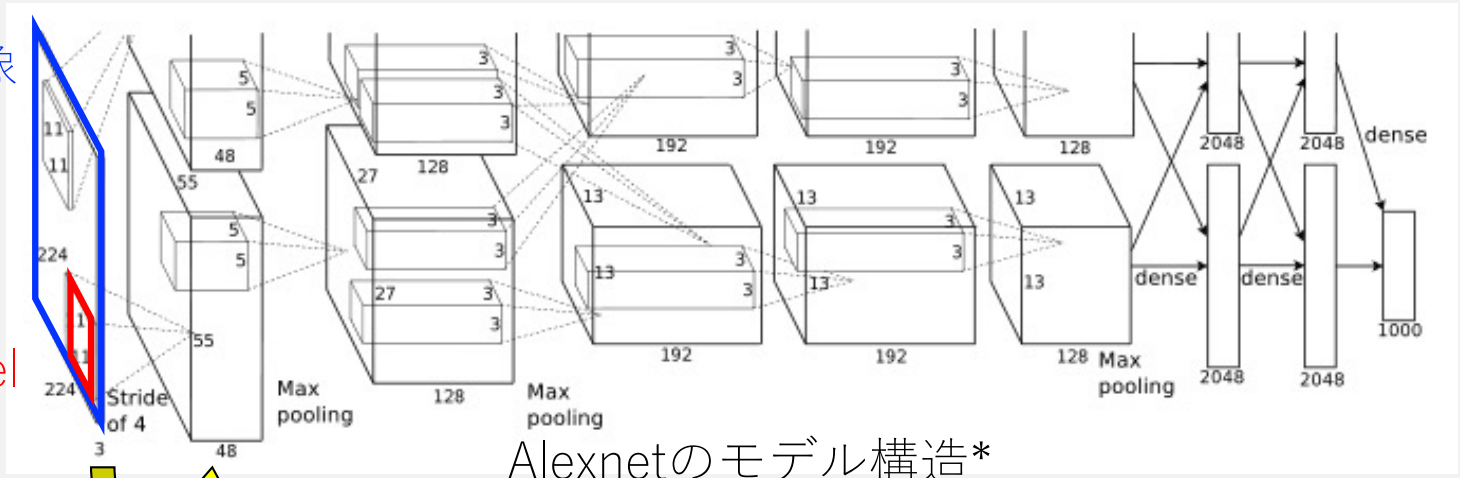
地盤調査データ（地質柱状図、密度検層、標準貫入試験、PS検層等）

- 国土交通省：Kunijiban
- 地盤工学会：新・関東の地盤、九州地盤情報共有データベース
- 防災科学技術研究所：ジオステーション
- 東京都土木技術支援・人材育成センター：東京の地盤

- 上記以外にも様々な機関でデータ公開がされている
- 本論文では、基盤深度のみを推定値としているが、上記データから評価できる様々な指標値（AVS30，地盤固有周期）に対してモデル構築可能

A.2.2 特徴量作成（画像における特徴抽出）

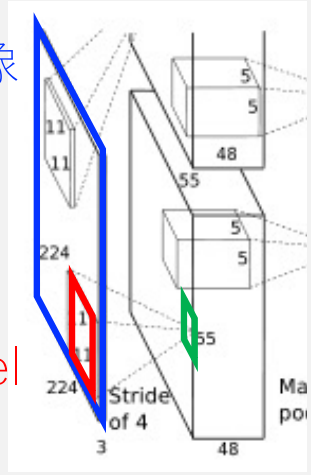
- 特徴量：入力データの特徴を表す数値
- 入力データ≠特徴量（例：犬、猫）
- 画像モデルではモデル内部で特徴抽出

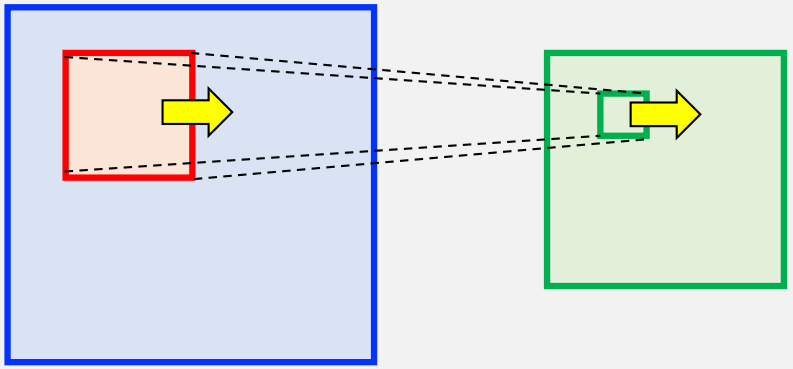


Alexnetのモデル構造*

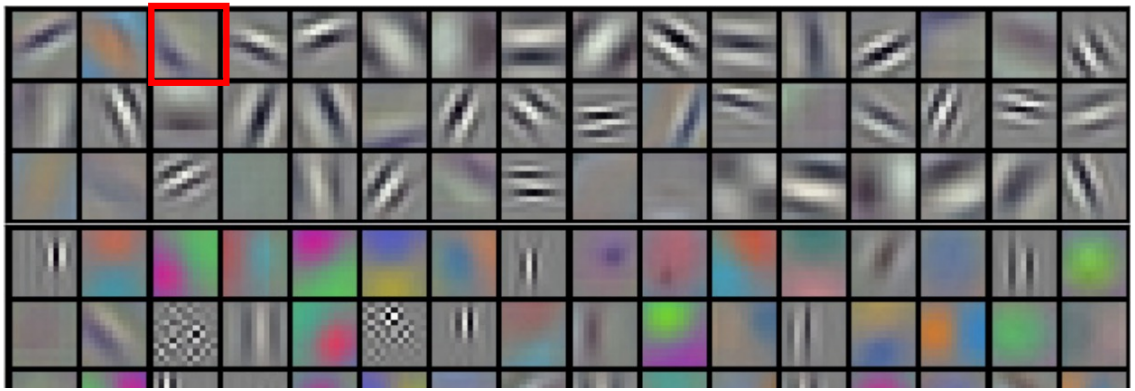
* ImageNet Classification with Deep Convolutional Neural Networks

A.2.2 特徴量作成（画像における特徴抽出）





Convolutionのイメージ

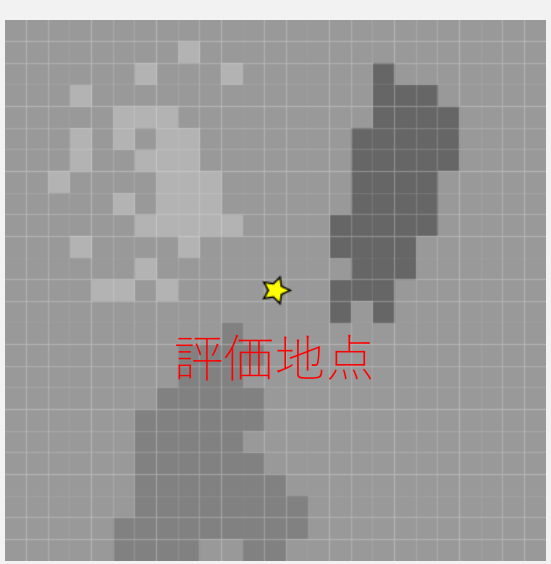


Alexnetの初期層における96個のkernelの重み*

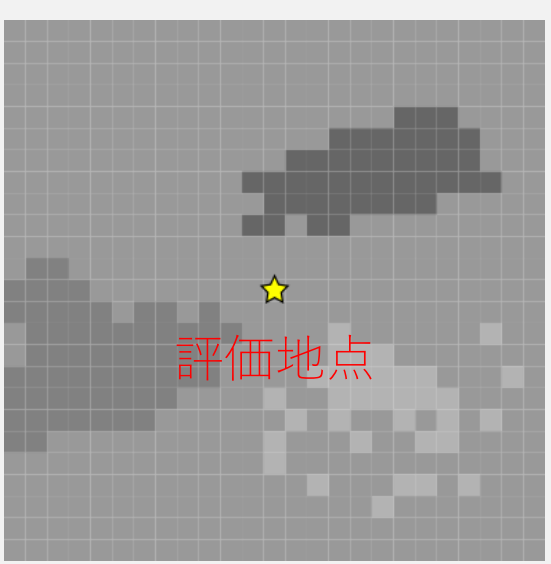
* ImageNet Classification with Deep Convolutional Neural Networks

A.2.2 特徴量作成（表形式データ）

- 特徴抽出のための特徴量エンジニアリングが必須
  - ➢ メッシュ値をそのまま入力とするのは不適切（例：回転・反転）



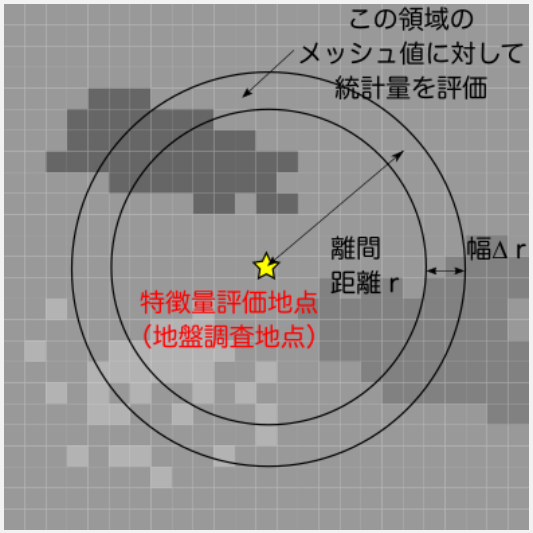
90度回転



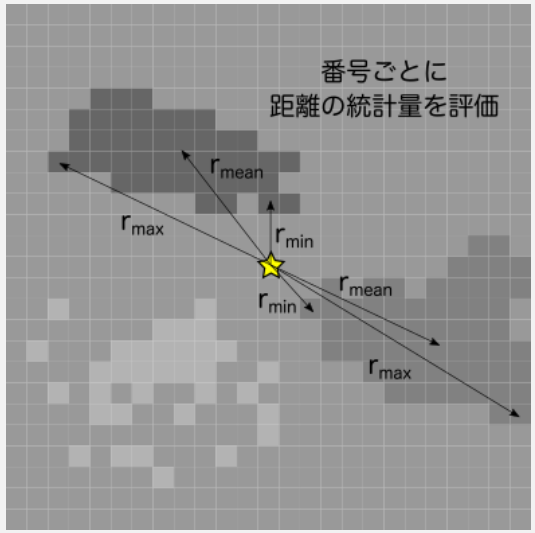
左右反転

A.2.2 特徴量作成（表形式データ）

- 特徴抽出のための特徴量エンジニアリングが必須
  - ➢ メッシュ値をそのまま入力とするのは不適切（例：回転）
  - ➢ これまでの専門知識が生きる部分でもある



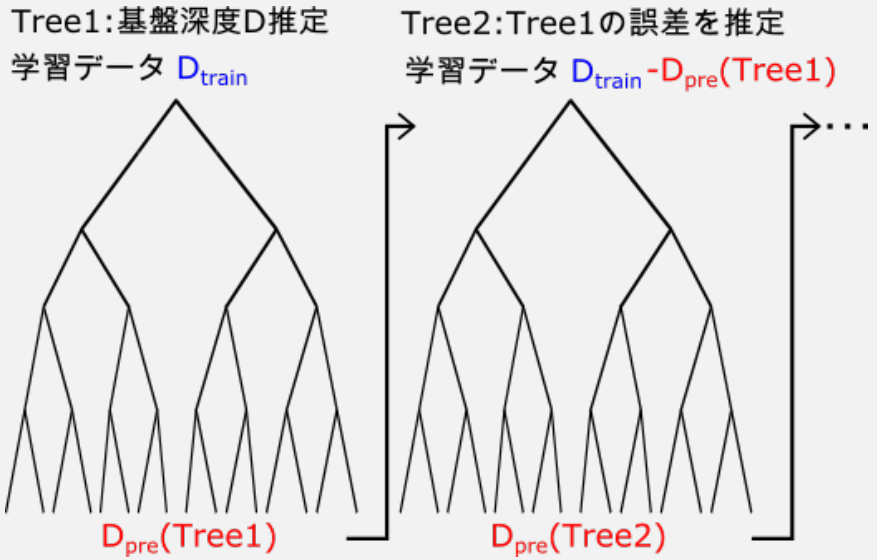
領域の平均値：周辺の状況をざっくり把握
領域の標準偏差：地形の変動状況を把握
（小さい：同じ地形が続く、大きい：異なる地形が含まれる）



地形分類番号（山地、台地・・・）に対する距離
r_{min}：その地形までの最小距離
r_{max}：その地形までの最大距離（拡がり）

A.2.3 モデルの作成

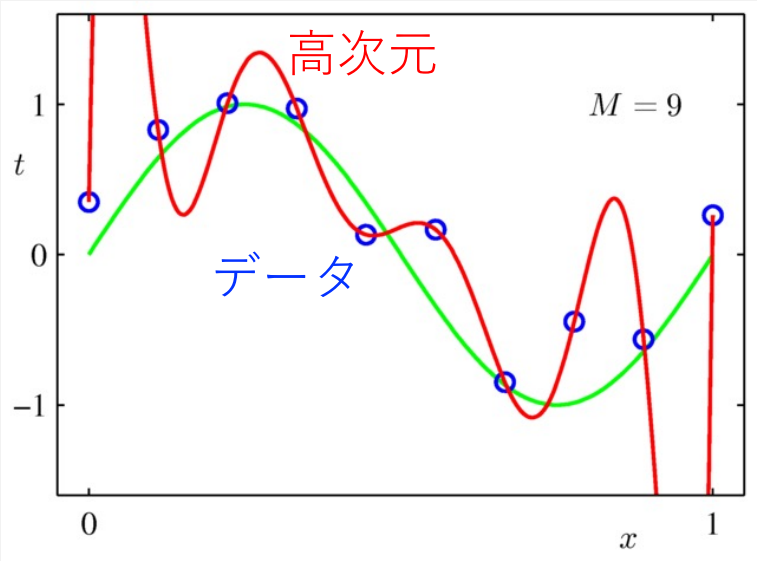
- Gradient Boosting Decision Tree（GBDT，LightGBM）
  - 特徴量の値に応じた条件分岐により、推定値を得る
  - 異なる決定木を直列化することで誤差低減する
  - ⇔Random Forest：異なる決定木を並列化
  - 過学習に注意（交差検証、正則化）



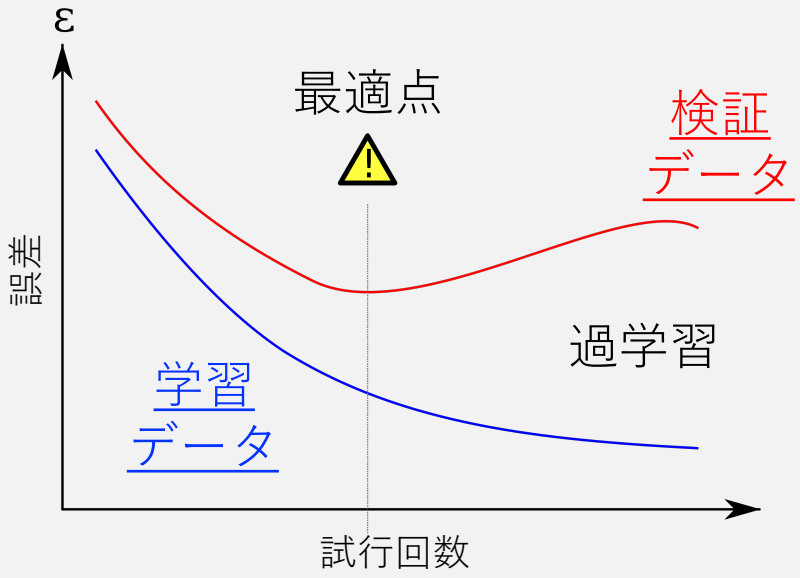
https://www.gormanalysis.com/blog/gradient-boosting-explained/

A.2.4 モデルの評価

- 過学習を防ぐための学習スキームの設計
  - データをモデル学習に使用する学習データとモデル性能を評価するための検証データに分割
  - 各データの誤差状況から学習回数の最適点を発見



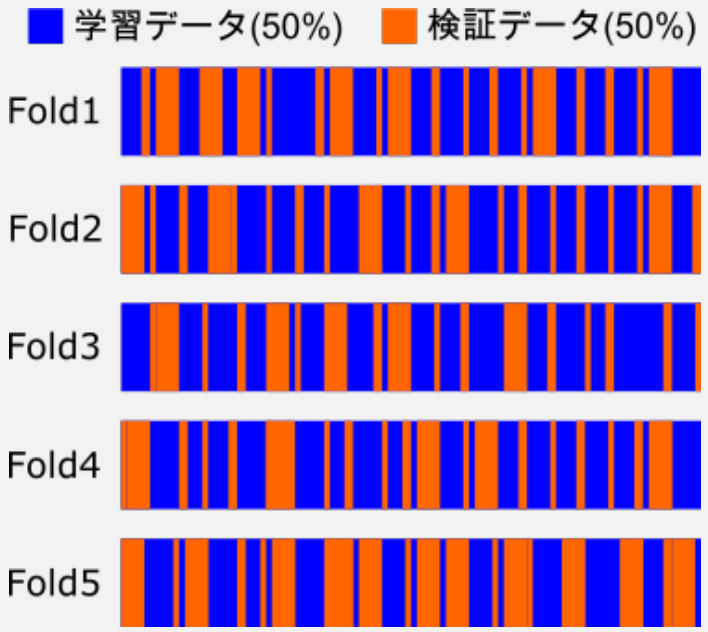
PRML(Fig1.4):https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf



学習・検証データに対する誤差
(Wikipedia 過剰適合の項に加筆)

A.2.4 モデルの評価

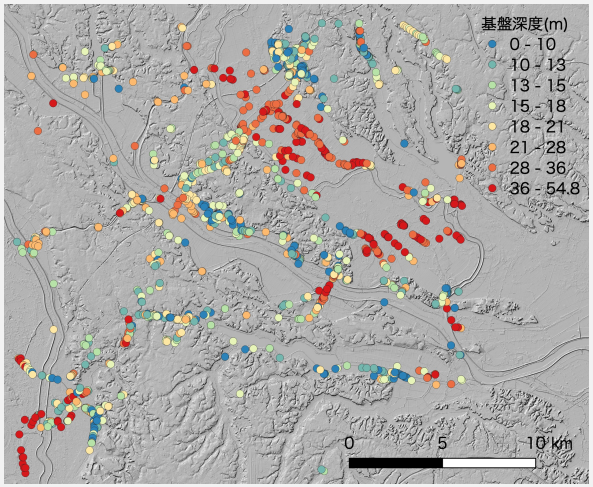
- 交差検証（Cross Validation）
  - ➢ 学習データと検証データの分割を変化させた複数Foldでモデルを構築
    - ✓ 全データを学習に使用するため
    - ✓ データセットの偏りによる推定結果の偏りを補正
- 本論文では空間的にランダムに散らばるようにデータ分割



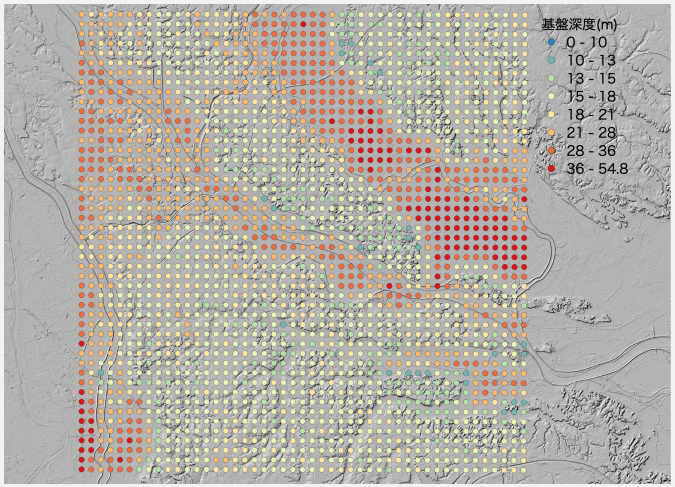
学習・検証データの分割

A.2.5 実サイトにおけるモデル構築

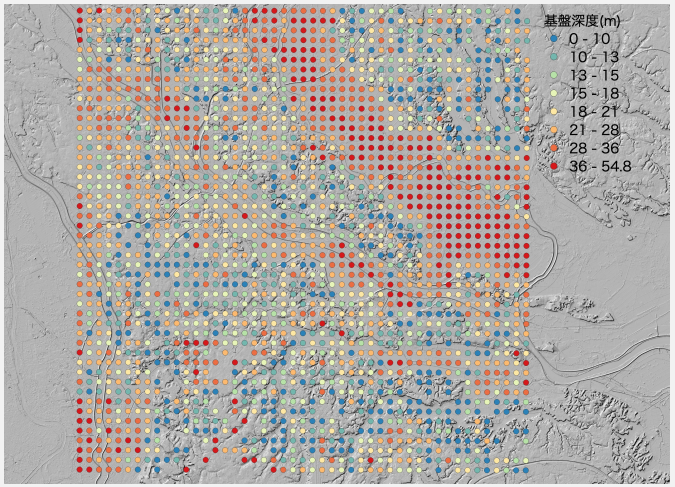
- 常総地域における結果
  - GBDTモデルではデータのない領域も安定した評価
  - Lasso回帰モデルでは不自然な変動（学習データに忠実）



モデル構築に使用した
地盤データ



構築したGBDTモデル
による面的推定結果

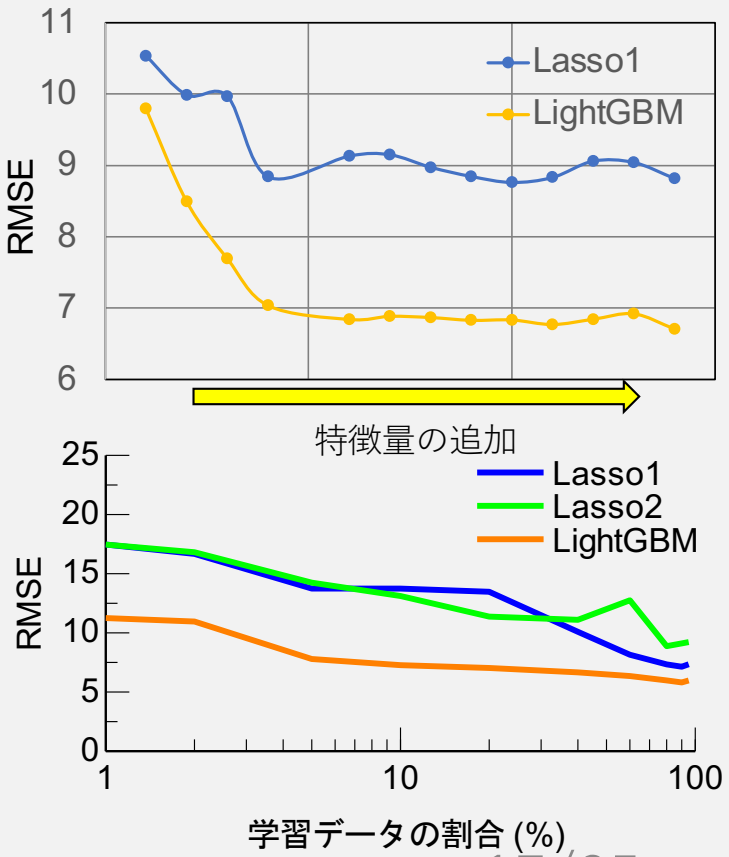


Lasso回帰モデル
による面的推定結果

A.2.5 実サイトにおけるモデル構築

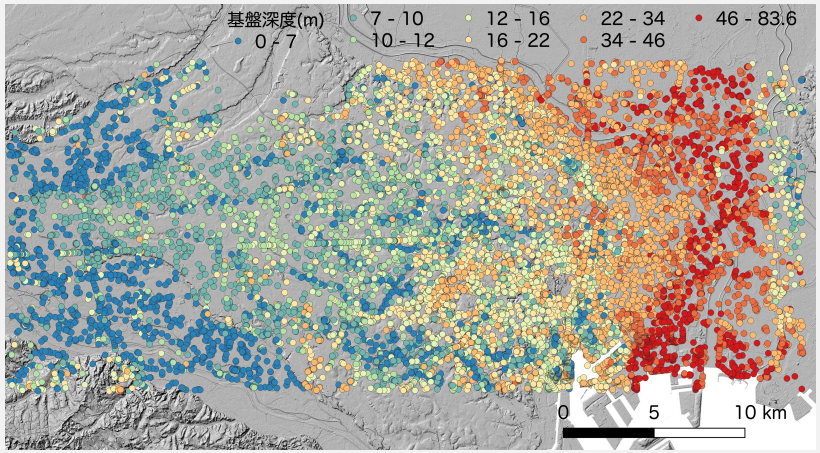
構築したモデルにおける各特徴量の重要度順位

順位	特徴量
1	最近傍ボーリングの基盤深度
2～6	離間距離100~1000mにあるボーリングの基盤深度の統計量
7	赤色立体地図の平均値の差分（離間距離100m－50m）
8	台地・段丘までの最小距離
9	氾濫平野までの最大距離
10	赤色立体地図の平均値の差分（離間距離300m－200m）
11	離間距離500~1000mにあるボーリングの基盤深度の統計量
12	評価地点の地形番号・地質番号に応じた基盤深度の平均値



A.2.5 実サイトにおけるモデル構築

- 東京地域における結果



モデル構築に使用した地盤データ

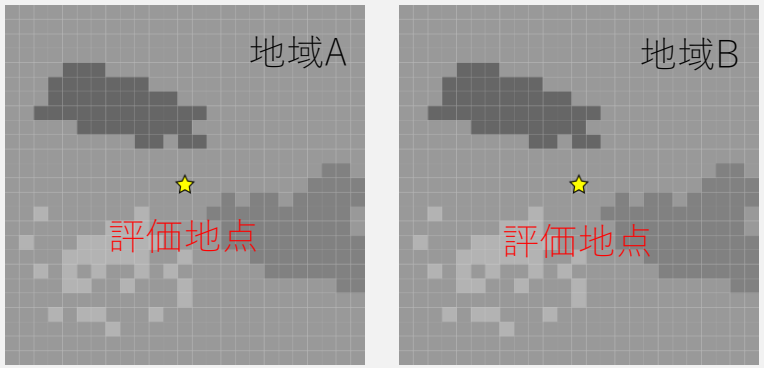
構築したGBDTモデルによる面的推定結果

A.2.6 結論

- 地形データから工学的基盤深度を推定するモデルを構築し、広域で作成した様々な特徴量を用いた機械学習モデルにおいて高い推定精度となることを確認した．一例ではあるが、地形データに対して機械学習モデルを構築することの有効性を確認した
- 従来の統計モデル（Lasso）と比較して機械学習モデルによる推定が優れている点を整理
  - 少ない学習データであっても高いモデル精度を発揮する
  - 学習データに対する過学習の可能性が低く、未知のデータに対する汎化性能が高い
  - 入力データの欠損値に関わらず、データの多寡に応じてモデルを使用できる

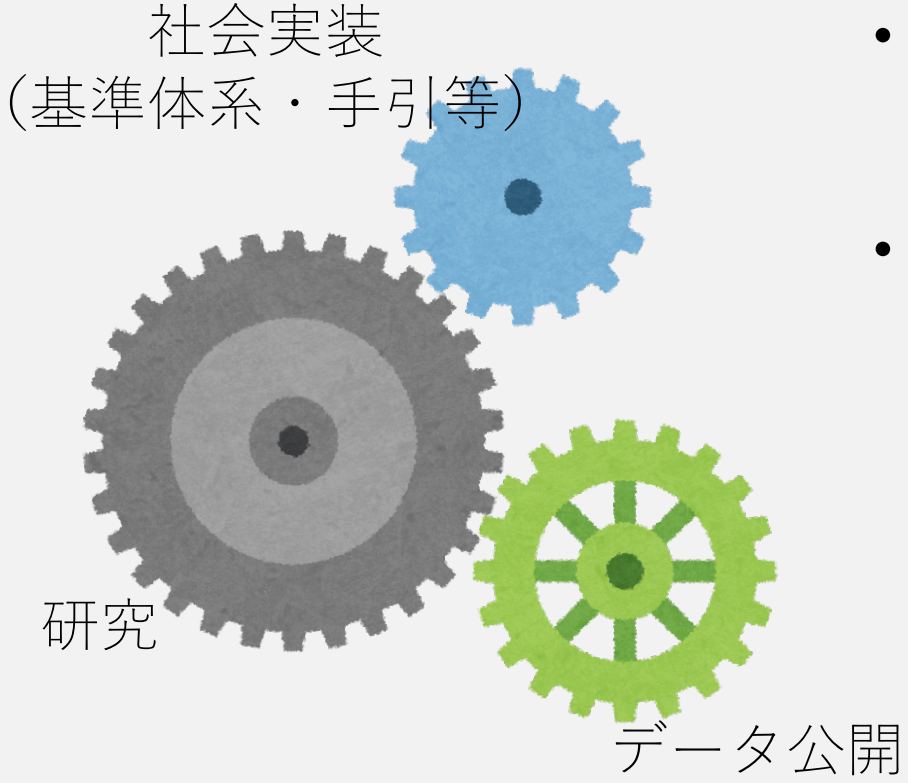
A.2.7 今後の課題・その他

- 本論文で有効性を確認した機械学習モデルを基盤深度以外の指標を推定するモデル構築に展開し，地形情報を影響要因とする様々な指標に対して分野横断的にモデルが構築できることを検証する（手元レベルでは確認済）
- 全国を対象とした大規模なモデルに拡張した場合についても考察を行う．地形情報だけでなく、地域性を反映できるようなモデルを構築する必要がある



- 本論文で使用したモデルはあくまで一例（Random Forestで同程度の精度を確認）であり、モデル構築のための考え方や手順を整備したという位置付けが強い。

B.1 地震工学の個人的な現状認識



- 多大な労力と予算で、３つが高いレベルで維持されてきたのが日本の地震工学
- 担い手や予算が減少する中でこのレベルが維持できるか？
  - 難しいのでは？
  - 個人的な認識では、技術者・設計者の負担大
  - 社会実装がボトルネックとなり、効果が見えにくい研究やデータ公開がシュリンクするという悪循環

B.2 技術者・設計者の負担を減らすためには

- 仕事をまとめる（例：手計算→電卓）
  - 仕事の具体に関わらず、高い抽象度で処理をまとめることができるのが機械学習（これができるならあれもできる）
    - 物体検出（Object Detection）は、自動運転、医療診断、衛星写真からの被災現場特定など、対象の具体に関わらず、各分野の専門家が行ってきた処理を代替できる。
    - 受賞論文でも、地形データを活用した予測技術に機械学習的なアプローチが活用できる可能性を示した。
- 武器を用意する
  - 成果のパッケージ化が容易
    - 文書（論文）だけではなくアルゴリズム（プログラム）
    - OSSの精神（知見の集約、エコシステムの醸成、プラットフォーム化による国際的なプレゼンス向上）

B.3 活用における課題（研究・実用）

- 再現性がない？　→　ある
  - 既往の論文と同様に条件を明記すれば再現は可能
  - データセット、ソースコードの公開によりこれまで以上の保証が可能
- 回帰分析と同じ？　→　違う
  - タスクの抽象化の程度が違う
  - タスクが同じであったとしても、圧倒的な性能によりゲームチェンジが起こる（AlphaGo，GPT3，AlphaFold2）
- ツールの応用は「研究」でない？　→　そうかもしれない。で？
  - 社会的ニーズが後押しする
  - 新たな場の生成（構造工学でのAI活用に関する研究小委員会の活動）
- ブラックボックスだから使えない？　→　これまでも同じ
  - 結果の判断理由が明確でない
  - 入力値によっては不安定な挙動をする可能性がある

おわりに

A. 受賞論文について

B. 今後の地震工学における機械学習の活用について

ご意見・コメントをお気軽にいただけると幸いです.

tanaka.kohei.22@rtri.or.jp