

機械学習による土質判定自動化の 可能性について

拓北地下開発株式会社 技術部

伊藤 正樹

自己紹介

- 最初はシステムエンジニア12年
 - 某専門学校情報システム科講師8年
 - 昨年からは地質調査の会社に入社
- 新たに始めた地質調査の業務に、これまでの自分のIT経験を生かして何か出来ることがあるのではないかと考えていました。

きっかけ

- それはコアチェックの時
- 最初に土質名を決めるのは、経験者の目と指先の感覚、時には匂いも。

→ 最初のコアチェックを、なるべく手間をかけずに何らかの方法で自動化できないか？

土質判定の自動化

- いま流行りの機械学習の手法を用いる
- 判定の材料としては、コア写真を使用
- 将来的には、粒度試験等の結果も加味



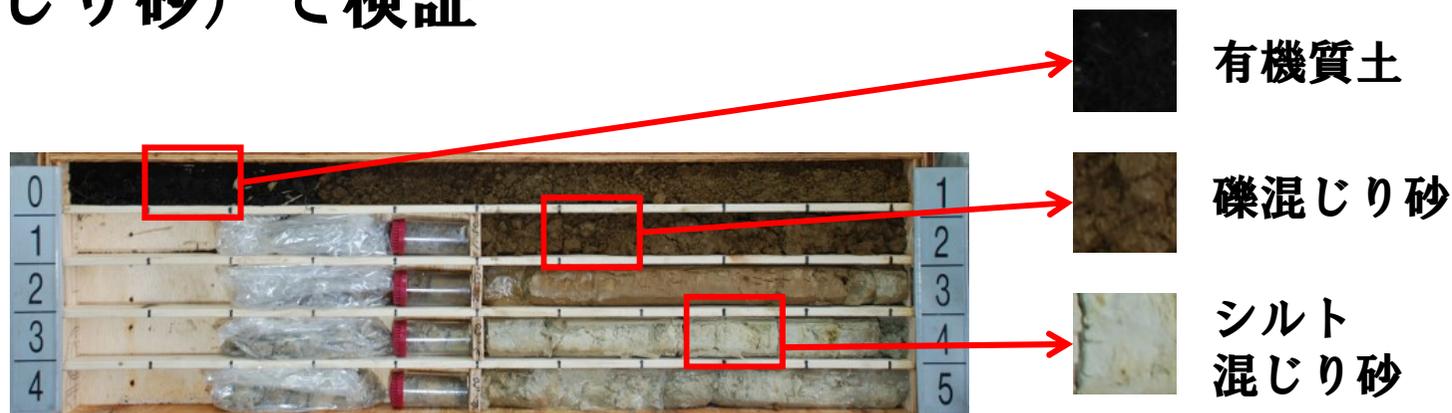
乱れた試料の
写真を使用

手順

1. コア写真を集め、データ化する
2. モデルを作成 (Octaveでプログラミング)
3. 2を1で繰り返し学習 (トレーニング)
4. 同じデータで正解率を計算してモデルを評価
5. 4の結果でパラメタチューニングや繰り返し回数の変更等を行う。2に戻る。
6. 5の結果で高い正解率が出たところで、未知のデータを用いテストを行う。

コア写真のデータ化

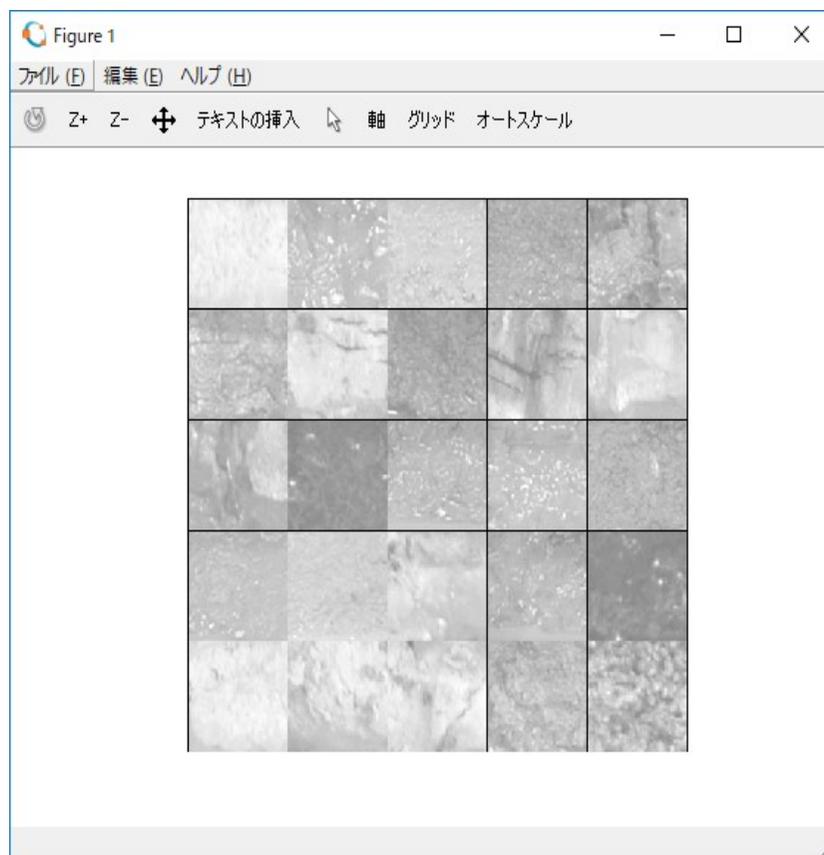
- コア写真から100×100ピクセルで画像を切り出し、土質名ラベルを付ける（フォルダに分類）
- 土質名は3つ（有機質土、礫混じり砂、シルト混じり砂）で検証



- データはトレーニング用とテスト用に分けておく

グレースケール化

- カラーではデータ量が多すぎるため、前処理として、グレースケール（白黒）化する。

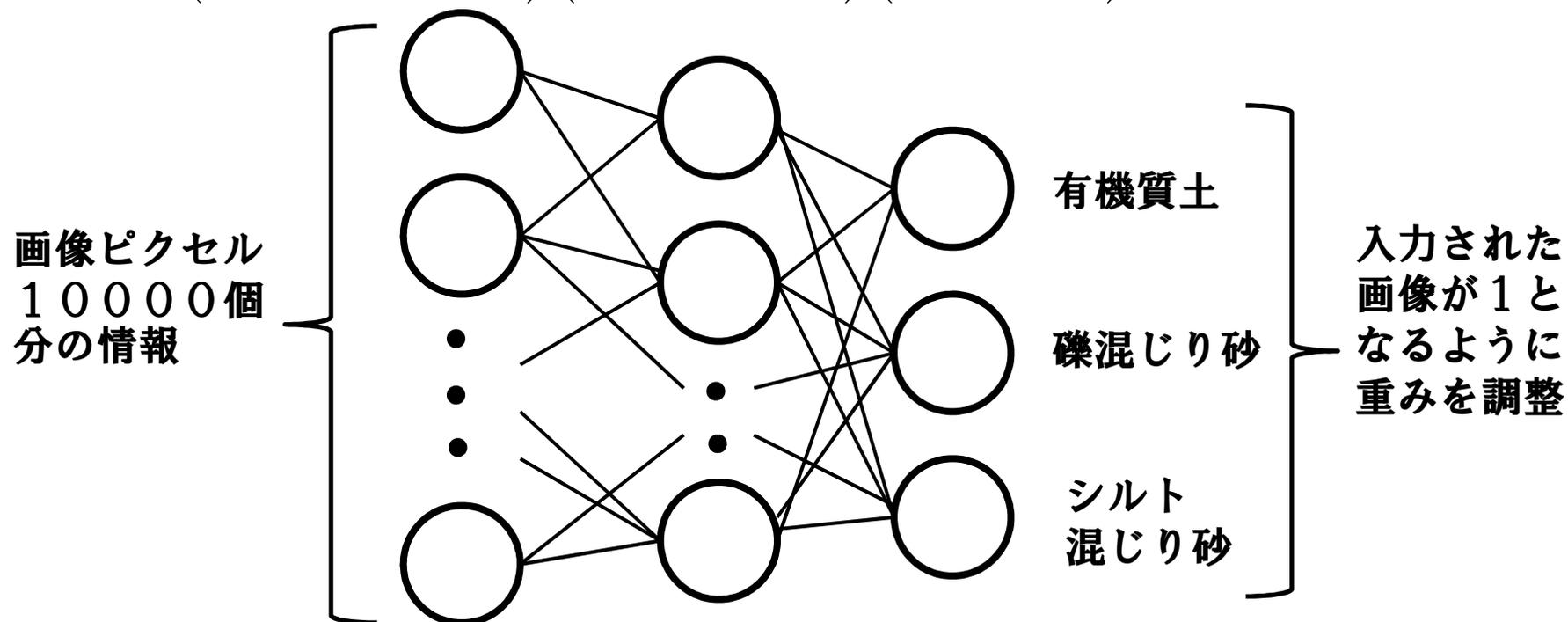


モデルを作成

- モデルにはニューラルネットワークを採用

入力層 中間層 出力層

(10000ノード) (100ノード) (3ノード)



モデルトレーニング中

The screenshot displays two windows from a Windows operating system. On the left is the Task Manager window, showing system performance metrics. On the right is the Octave software window, showing the progress of a neural network training process.

Task Manager (タスクマネージャー) Performance Tab:

- CPU:** Intel(R) Celeron(R) CPU N3050@ 1.60GHz, Usage: 100%, 2.15 GHz
- メモリ:** 2.5/7.9 GB (32%)
- ディスク 0 (C:):** 0%
- イーサネット:** 送信: 0 受信: 0
- Wi-Fi:** 接続されています
- Bluetooth:** 接続されています

Octave Command Window Output:

```
Loading and Visualizing Data ...
Program paused. Press enter to continue.

Initializing Neural Network Parameters ...

Training Neural Network...
Iteration 70 | Cost: 1.795891e+000
```

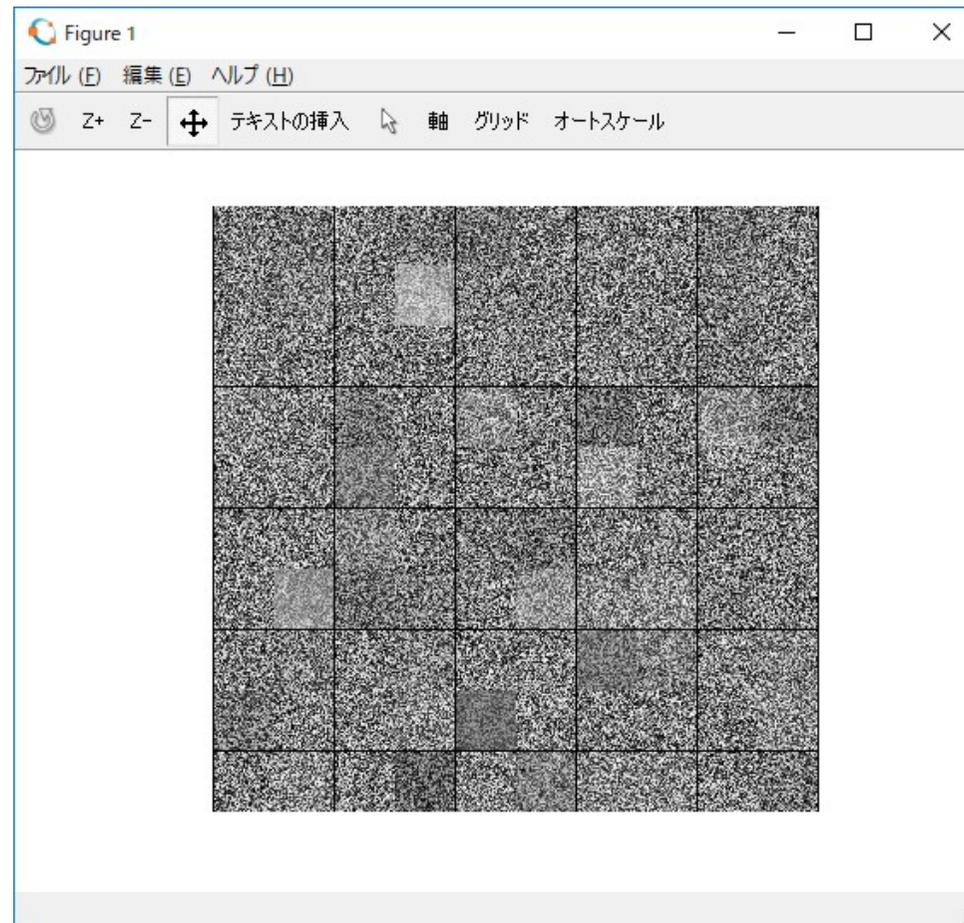
The Task Manager window also displays a graph of CPU usage over 60 seconds, showing a spike to 100%. Below the graph, the following system information is shown:

項目	値	項目	値
使用率	100%	最大速度	1.60 GHz
速度	2.15 GHz	ソケット	1
プロセス数	151	コア	2
スレッド数	1405	論理プロセッサ数	2
ハンドル数	57236	仮想化	無効
稼働時間	4:16:27:39	Hyper-V サポート	はい
		L1 キャッシュ	112 KB
		L2 キャッシュ	2.0 MB

The Octave window shows a file browser on the left with a directory structure including 'img', 'imgnew', 'imgorg', 'lib', and '.DS_Store'. The command history on the right shows the execution of 'pkg load image' and the current iteration status.

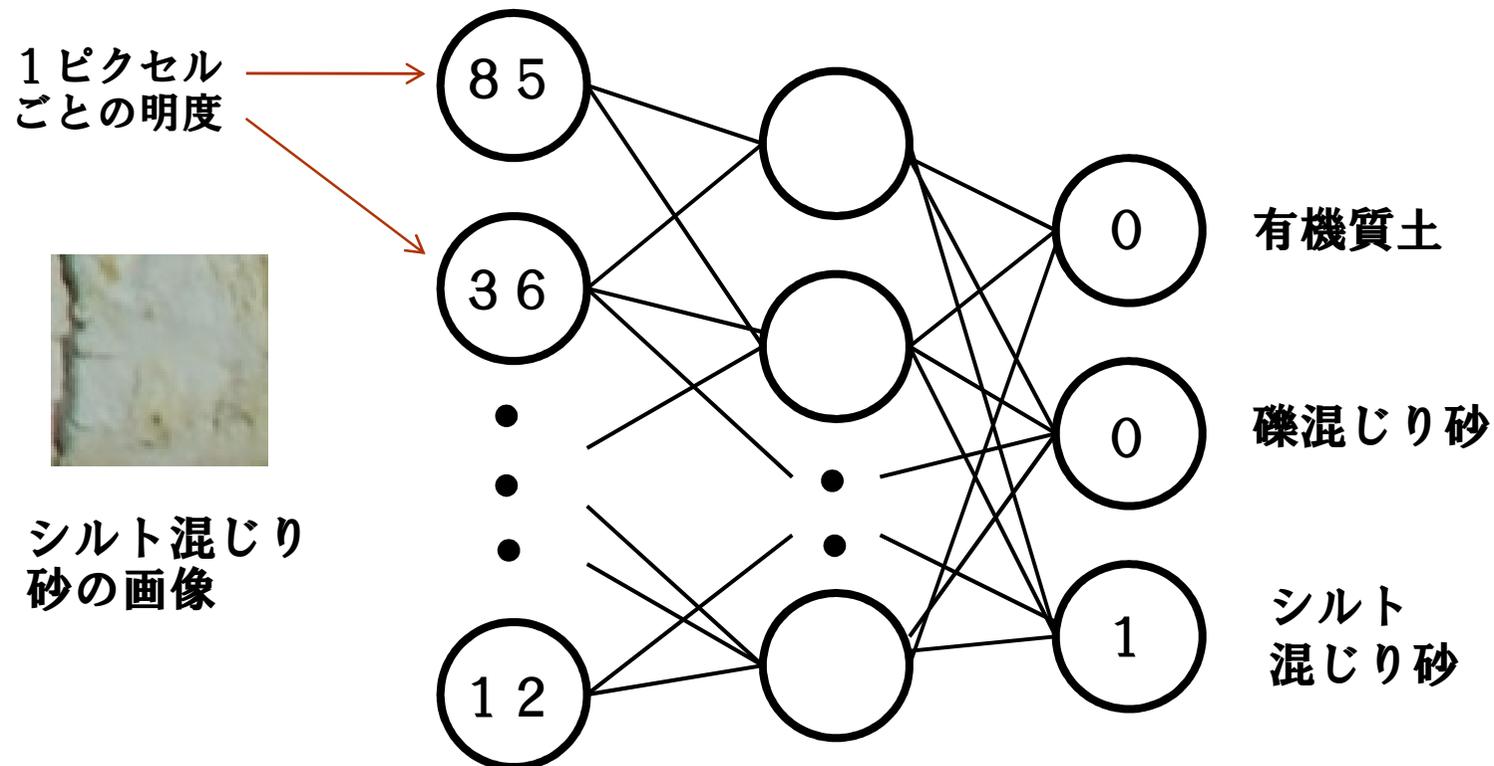
中間層イメージ

- トレーニング後の中間層の状態イメージ



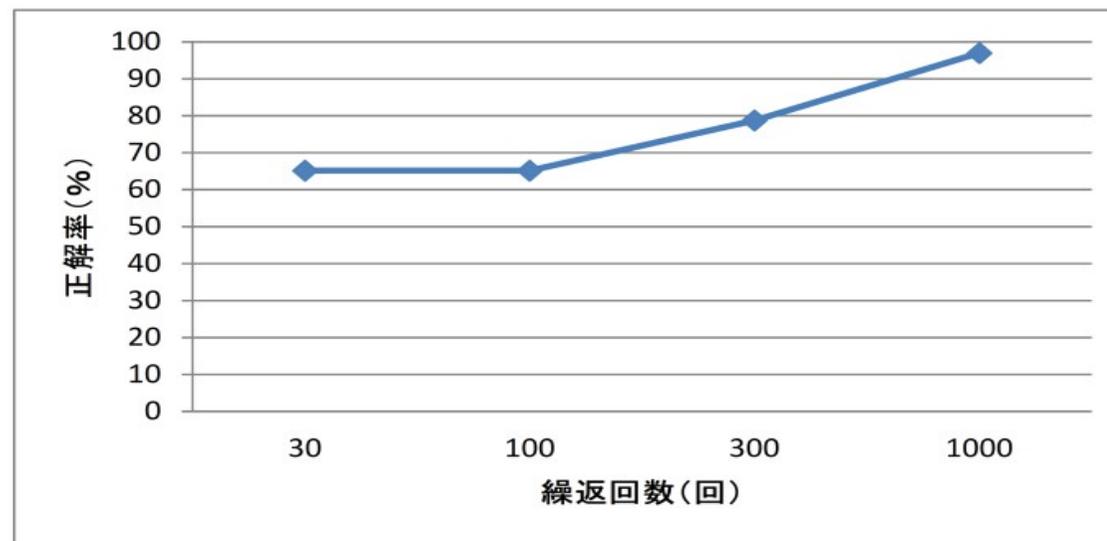
判定イメージ

トレーニング後は、シルト混じり砂の画像入力すると、出力層はシルト混じり砂1、それ以外0となる。



トレーニングデータ正解率確認

- トレーニングの繰り返し回数を変えて、トレーニング用データに対してどこまで正解率が上がるかをグラフ化した。
- 繰り返し回数を増やせば増やすほど正解率が上がる傾向がみられた。（3回実施した平均値）



未知のデータで判定

- 次に成績の良かった学習済みモデルを用いて、判定を10回行った結果は以下。

礫混じり砂		→ 90%正解
シルト混じり砂		→ 100%正解
有機質土		→ 10回すべてで上記2つに誤判定 正解率0% (内シルト混じり砂と 認識されるケースが8回)

過学習の問題も考えられる

誤判定例

```
Octave
ファイル (E) 編集 (E) デバッグ (B) ウィンドウ (W) ヘルプ (H) コース (N)
現在のディレクトリ: C:\soil\soil
ファイルブラウザ
C:/soil/soil
名前
ex4data1.mat
ex4weights.mat
fmincg.m
nnCostFunction.m
octave-workspace
predict.m
randInitializeWeights.m
sigmoid.m
sigmoidGradient.m
ワークスペース
名前 / クラス 次元
A uint8 100x100x3
B double 100x100
Imgfilename char 1x21
Theta1 double 100x10001
Theta2 double 3x101
X double 155x10000
XTEST double 3x10000
コマンドウィンドウ
Loading and Visualizing Data ...
Program paused. Press enter to continue.
Initializing Neural Network Parameters ...
Training Neural Network...
Iteration 1000 | Cost: 7.494343e-001
Program paused. Press enter to continue.
Visualizing Neural Network...
Program paused. Press enter to continue.
Training Set Accuracy: 90.967742
Testpred =
1
2
1
Test Set Accuracy: 66.666667
>>
コマンド履歴
ex4
ex4
ex4
ex4
ex4
ex4
```

本来であれば、
1 (礫混じり砂)
2 (シルト混じり砂)
3 (有機質土)
と表示されていなければならないところが、
1 (礫混じり砂)
2 (シルト混じり砂)
1 (礫混じり砂)
と判定されている。

今後の課題

- **過学習の問題**（トレーニング用データに対しては90%以上の高い正解率であっても、未知のテストデータに関しては低い）
 - データを多くする、ネットワーク構成を変える、パラメタを調整する等が考えられる。
- **画像にコントラストが無いもの（黒一色）は識別が難しい**
 - コントラスト補正（正規化）や、色情報も加えることも考慮

実験を通して感じたこと

- **機械学習は万能ではない**

モデルのノードの数、中間層の数をどうするか、パラメタのチューニングは人の手がかかる。うまくいかない時の解決法までは教えてくれない。

- **データの前処理にも労力が必要**

写真の切り出し、分類してラベル付け、大きさをそろえる等、トレーニングを実際にさせる以上に前準備に時間がかかる。

詳しい内容

- 今回の発表の詳しい内容は、弊社HP内で紹介しています。

<http://www.takuhoku-chika.co.jp/autohantei.html>

ご清聴ありがとうございました