

機械学習を用いたシールドマシン操作判定モデルにおけるデータサンプリングの影響確認

清水建設(株) 正会員 ○鎌田浩基, 陳劍, 高本尚彦, 杉山博一, 山本真哉, 野澤剛二郎, 本多眞

1. はじめに

著者らは、AIによるシールドマシンの自動運転を目指したAIモデルの開発の一環として、シールドマシンのジャッキ操作を学習したAIモデル(操作支援AIモデル)の開発を進めてきた。操作支援AIモデルは、ジャッキ操作のタイミングを判定するAIモデル(以下、操作判定AIモデル)と設定値を判定するAIモデル(以下、操作予測AIモデル)の二つからなり、操作タイミングを考慮した実際のオペレータに近い予測が可能なモデルである。また、シールドマシンのジャッキ操作は、操作変更の頻度が数十秒から数百秒に一回程度であることに着目し、操作支援AIモデルではそのデータバランスを調整するためにデータの重みづけを行っている。しかし、操作判定モデルでは学習データ内において操作変更の回数が著しく少ない場合に、予測モデルの精度が低下してしまうことが分かってきた。本稿では、操作支援AIにおける、操作判定AIモデルの精度向上のために、データサンプリングの効果確認を実施した。

2. 操作判定AIモデルの構築

操作判定AIモデルの学習方法について図-1に示す。はじめに、①操作しなかった記録(操作無記録)を0、オペレータが操作したときの記録(操作有記録)を1として分類する。②操作無記録の総数 n_0 と操作有記録の総数 n_1 のデータ数の偏りを考慮し、データの重みづけを行う。ここでは、操作無記録の重みを1、操作有記録の重みを n_0/n_1 とした。次に、③手順②で定めた重みを考慮し、操作有無判定値を目的変数とする教師有学習により操作判定AIモデルを作成する。本稿では、勾配ブースティング決定木により操作判定AIモデルの学習を実施した。④学習した操作判定AIモデルによる予測値により確率分布を求め、TN、FN、TP、FPのデータ数が適切になるように閾値を決定する。本稿では、下記に示すBER(Balanced Error Rate)が最小になるよう設定した。

$$BER = 0.5 \times \left(\frac{FP}{TN + FP} + \frac{FN}{FN + TP} \right) \quad (1)$$

3. 操作予測AIモデルの構築

操作予測AIモデルは、シールドジャッキ操作の設定値を目的変数とする教師あり学習により作成する。その際、教師データは操作判定AIモデルにより操作有と予測された記録、すなわちTPとFNのみを用いた。これは、操作変更が必要であると予測されたデータのみを用いることにより、学習に用いる記録の操作バランスを整えるためである。これにより、常に操作しないことを許容するAIモデルとなることを防ぐことが可能である。本稿では、勾配ブースティング回帰木を用いて学習を行った。

4. 操作判定AIモデルにおけるデータサンプリング手法

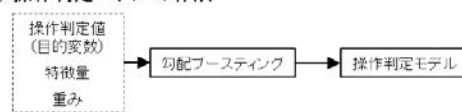
オペレータのジャッキ操作回数は、掘進する地質状況や線形により大きく変化することがわかっている。掘進中に操作回数が特に少ない場合には、操作判定AIにおいて十分な学習を行うことができない。

① 操作有無の判定

② データの重みづけ

操作無記録			操作有記録		
レコード番号	操作判定値	重み	レコード番号	操作判定値	重み
1	0	1	1	1	n_0/n_1
2	0	1	2	1	n_0/n_1
⋮	⋮	⋮	⋮	⋮	⋮
n_0	0	1	n_1	1	n_0/n_1

③ 操作判定モデルの作成



④ 閾値の決定

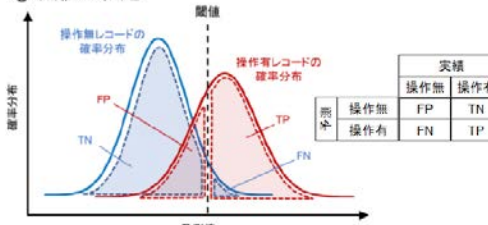


図-1 操作判定モデルの作成手順

キーワード シールドトンネル, 人工知能, 自動操縦, データサンプリング

連絡先 〒135-8530 東京都江東区越中島 3-4-17 清水建設株式会社 技術研究所 TEL 03-3820-6891

その一方でオペレータが実際にジャッキ操作をするのは一瞬だが、操作時と同じような状況はある程度時間的な幅をもつことが工学的に予想される。そこで本稿では、操作有レコードの n_{smp} 秒前以内のデータであるならば、操作無レコードであっても操作有レコードへと変更(サンプリング)可能であると仮定して学習を行い、操作有無モデルの予測性能の向上を図る。

データサンプリングに用いる秒数 n_{smp} には、0秒(データサンプリングなし)、5秒、10秒、20秒、30秒のそれぞれを設定しその結果の比較を行う。また、操作判定AIモデルの評価は、実際の操作タイミングが30秒の幅があると仮定し、30秒のサンプリングを施したホールドアウトデータを用い、荷重 Logloss と適合率(Precision)、再現率(Recall)の三種類の指標を用いて行った。なお、修正再現率(Modified recall)については、(正解した操作有予測数、 n_{smp} 秒以内で重複カウントなし)/(サンプリングしない場合の操作有レコード数)として計算した。

$$\text{Logloss} = -\{y \log p + (1 - y) \log(1 - p)\} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

5. 操作判定AIモデルにおけるデータサンプリング検証結果

表1に検証結果を示す。荷重 Logloss をみると、サンプリング秒数 n_{smp} を増やすに従い、モデルの性能が高まることがわかる。モデル評価は30秒のデータサンプリングを施したデータを用いているため、この結果は妥当であると考えられる。また、適合率についてみると、サンプリングによるモデルの改良は必ずしもみられないことがわかる。このように低い適合率は、操作無レコードを操作有と予測していることを表しているが、操作が必要なタイミングが操作無レコードでも存在していることが原因であると考えられる。このようなデータをさらに効果的に学習するためには、クラスタリングなどによるデータサンプリングが有効であると考えられる。次に再現率についてみると、 n_{smp} を10秒としたモデルにピークがあり、いずれもサンプリングを行わないものより値が高いことがわかる。ゆえにモデル評価時と異なる n_{smp} であっても提案するサンプリングの実施が有効であることがわかる。この結果は、荷重 Logloss における傾向と異なるが、これは Logloss が、0または1に近い操作確率を予測するモデルほど性能を高く評価する性質をもつことが要因の一つと考えられる。また、サンプリングをしない操作回数に基づいた、修正再現率をみると、0秒のサンプリングに対し、5秒～30秒のサンプリングではモデル性能が高まっていることがわかる。ただし、そのピークは不明瞭である。

表1 データサンプリング検証結果

	0s	5s	10s	20s	30s
Weighted logloss	1.115	0.9535	0.8407	0.7832	0.7027
Recall	0.219	0.241	0.370	0.244	0.296
Recall(modified)	0.455	0.545	0.727	0.545	0.818
Precision	0.019	0.017	0.024	0.013	0.017

6. まとめ

本稿では、操作判定AIモデル構築におけるデータサンプリング手法の提案とその効果について確認した。今回は単純化のため、データサンプリングの方法はジャッキ操作時から一律に n_{smp} 秒前までの範囲で行ったが、ジャッキ操作のタイミングごとに異なる秒数でのサンプリングを行うべきと考えられるため、今後はクラスタリングなどを用いたデータサンプリングを行いたいと考えている。

参考文献

- 1) 和田健介, 杉山博一, 野澤剛二郎, 本多眞: シールドマシンの操作特性を考慮した自動操縦AIモデル, 令和元年度土木学会第74回年次学術講演会, VI-814, 2019