

潤滑油の耐摩耗性能予測における機械学習の適用

Application of Machine Learning in Anti-wear Performance Prediction of Lubricants

出光興産(正) *小林 兼士

Kenji Kobayashi

Lubricants Research Laboratory, Idemitsu Kosan Co, Ltd., Japan

1. 目的

既存のデータを数理モデルにより学習し、一定のルールやパターンを発見する方法である機械学習は、学習結果を基に、未知の結果の予測に用いることができる。本報告の目的は、潤滑油を構成する原材料の配合量から、耐摩耗性能の予測を行った結果を基に、潤滑油の性能予測に適した数理モデルについて考察することである。

2. 検証方法

同一条件で測定した四球摩耗試験(ASTM D4172 準拠)の試験結果(摩耗痕幅, mm)を 741 個収集し検証に用いた。供試した潤滑油の原材料配合量を説明変数、試験結果を目的変数とするデータベースを作成した。データベースの模式的なイメージを図 1 に示す。試験結果は、0.50mm 以下の結果を「1」、0.50mm を上回る結果を「0」と二値化した。

741 個の試験結果から、20 個の試験結果を抽出し、これを検証用データとした。そして残り 721 個のデータを学習用データとした。各予測手法にて学習結果に基づく予測値と実測値を照合し、その正答率を比較した。予測手法として、ニューラルネットワークモデルを用いた機械学習と、多変量解析として一般的な重回帰あるいはロジスティック重回帰を用いた手法を選択し、3 つの手法の予測精度を比較した。

ニューラルネットワークを用いた機械学習を行うにあたり、三つの全結合層を中間層に持つモデルを構築し、学習を行った。モデルの模式的なイメージを図 2 に示す。各ノードは入力された情報を変換する役割を持ち、変換に用いる活性化関数は ReLU (Rectified linear function, 正規化線形関数)とした。ただし、予測結果を出力する出力層の活性化関数のみ、シグモイド関数とした。その他の各層のノード数や、一度に処理するデータ数(ミニバッチサイズ)など、予め設定する必要がある変数(ハイパーパラメータ)は、十分な学習率が得られるよう試行錯誤により決定した。学習は乱数を用いるため、同じ学習データを用いても試行毎に学習結果が異なることが想定される。そこで学習は異なる乱数設定で 3 回行い、それぞれの学習結果にて正答率を算出し、予想精度の妥当性を判断できるようにした。

重回帰による予測を行うにあたり、式(1)に示す回帰式を用いて学習を行った。目的変数(y)は、二値化する前の試験結果であり、各原材料の配合比を説明変数(x)とした。回帰式により得られる予測値と実測値との誤差が最も少なくなるよう説明変数の各係数(a)を最適化した。最適化された回帰式に検証用データの説明変数を代入し、得られた予測値を二値化して、実際の二値化した試験結果と照合することで、正答率を算出した。

ロジスティック重回帰による予測を行うにあたり、式(2)に示す回帰式を用いて学習を行った。ロジスティック重回帰は重回帰とは異なり、目的変数は二値化された試験結果である。式(2)の P は確率を示し、本検証においては P が 0.5 より大きければ、二値化された試験結果にて「0」を示す確率が、「1」を示す確率より高いと判断できる。回帰式により得られる予測値と実測値との誤差が、最も少なくなるよう説明変数(a)を最適化した。最適化された回帰式に検証用データの説明変数を代入し、得られた予測値を、実際の試験結果と照合することで、正答率を算出した。

	Material dosage, %				Test result	Test result (Binarization)
	Material A	Material B	...	Material ...		
Sample 1	60.0	3.0	...	0.2	0.68	0
Sample 2	40.0	5.0	...	0	0.47	1
...
Sample 721	20.0	0.0	...	0.1	0.55	0

Fig.1 Image of database

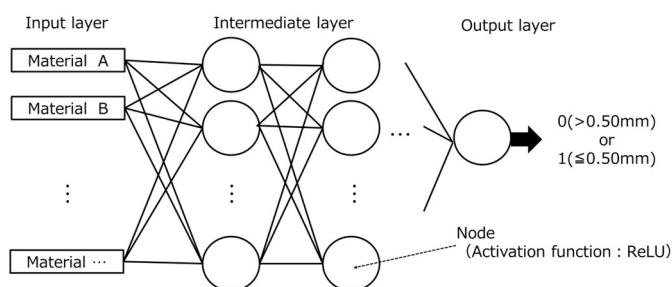


Fig.2 Image of neural network model

$$y = a_0 + a_A x_A + a_B x_B + \dots + a_{\dots} x_{\dots} \quad \dots (1)$$

$$P = \frac{1}{1 + e^{-(a_0 + a_A x_A + a_B x_B + \dots + a_{\dots} x_{\dots})}} \quad \dots (2)$$

3. 検証結果と考察

各手法における正答率を表 1 に示す。重回帰およびロジスティック重回帰を用いた場合、ランダムに回答した場合に期待される正答率(50%)と同等の正答率であり、十分な予測精度ではなかった。一方でニューラルネットワークモデル

を用いた機械学習では、70%から 80%の正答率が得られ、比較的良好な予測精度であった。

721 個の学習用データについて、重回帰にて得られた予測値と、実測値の関係を図 3 に示す。図 3 において決定係数 R^2 は 0.40 であり、十分に最適化ができていないと判断できる。また同じく学習用データについて、ロジスティック回帰により算出した二値化された予測値と、二値化した実測値の関係を表 2 に示す。学習用データにおける正答率は 85% であり、ロジスティック回帰は、学習データを比較的高い精度で予測できていることが分かる。しかし、学習用データにはない検証用データに対して、十分な予測精度を示さなかったのは、学習データへ過度に回帰式が適合することで、未知のデータを予測する性能(汎化性能)が低下する現象である、過学習(オーバーフィッティング)が生じたためと考えられる。

ニューラルネットワークモデルを用いた機械学習が、重回帰およびロジスティック回帰による予測より高い正答率を示した理由として、説明変数と目的変数の非線形な関係を学習できたことが可能性として考えられる。説明変数となる各原材料の配合量と、目的変数となる耐摩耗性能の関係は、必ずしも直線的な関係にあるとは限らない。例えば、耐摩耗性向上に効果を持つある原材料が、過配合による腐食摩耗で耐摩耗性を逆に悪化させるような事象があった場合、線形分析である重回帰およびロジスティック回帰を用いたモデルでは、この事象を学習することができない。一方で、ニューラルネットワークモデルはこのような非線形な関係を学習することができる。また例えば、複数の添加剤同士の交互作用も、非線形な関係となる場合がある。耐摩耗性向上機能を有する 2 つの添加剤が、併用により機能を失った場合を例に取り、原材料配合量と耐摩耗性の非線形な関係と、その結果に対するニューラルネットワークモデルによる学習イメージを図 4 に示す。図 4 は簡易化したモデルであるが、今回の検証において用いたデータにおいても、前述の例と類似するような非線形な関係が、原材料配合量と測定結果の間に含まれていた可能性がある。

Table 1 Correct answer rate at each predict method

	Multiple regression	Logistic regression	Machine learning with neural network model
Correct answer rate, %	45	50	70 ~ 80 (N1:75, N2:80, N3:70)

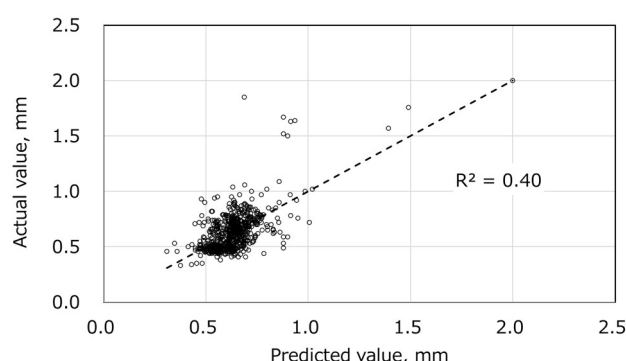
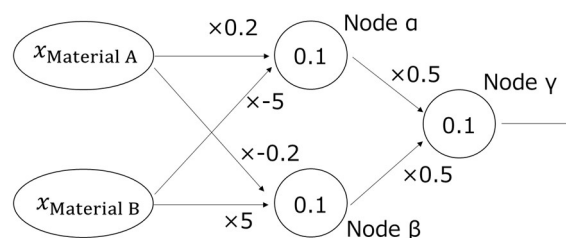


Fig.3 Actual and predicted value in multiple regression

Table 2 Actual and predicted value in logistic regression

Frequency		Predicted value	
		0	1
Actual value	0 (> 0.50)	486	47
	1 (≤ 0.50)	58	130



	Material A	Material B	Binary result	Output of each node		
	A	B		α	β	γ
SAMPLE I	0.0	0.0	0	0	0	0
SAMPLE II	2.5	0.0	1	1	0	1
SAMPLE III	0.0	0.1	1	0	1	1
SAMPLE IV	2.5	0.1	0	0	0	0

Activation function of each node : $y = \begin{cases} 1, (a_A x_A + a_B x_B \geq c) \\ 0, (a_A x_A + a_B x_B < c) \end{cases}$

a_A and a_B are optimized in each node.
 c is set to 0.1 in this example.

Fig.4 Simple example of learning interaction of materials and performance by neural network

4. まとめと今後の課題

本研究では、潤滑油配合量から耐摩耗性能を予測する検証を行い、重回帰やロジスティック回帰などの多変量解析を用いた予測よりも、ニューラルネットワークモデルを用いた予測が、予測精度に優れているという結果が得られた。この理由として、ニューラルネットワークモデルが、原材料配合量と耐摩耗性の非線形な関係を学習できたことが考えられる。耐摩耗性以外の性能においても、潤滑油を構成する原材料の配合量と性能との関係は、非線形な関係になる可能性がある。例えば摩擦係数の予測においては、重回帰を用いたモデルでは十分な予測精度が得られず、決定木を用いた機械学習であるランダムフォレストにより良好な精度が得られたという報告もある¹⁾。今後、原材料配合量と耐摩耗性以外の性能についても、同様の検証を行う予定である。

本研究のさらなる課題として、予測精度の向上が挙げられる。予測精度を向上させるためのアプローチとして、動粘度や元素分などの、性能との相関が期待される項目も学習させることや、ニューラルネットワークモデル以外の、非線形な関係を学習できるモデルを用いることなどが考えられる。

文献

- 1) 野間・青木・小林：異種添加剤併用系における鋼の摩擦係数予測に向けた機械学習の応用，トライボロジー会議 2022 秋 福井 予稿集, (2022) B29.