

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

Application of Machine Learning for Predicting Friction Coefficient of Steel under Combination of Multiple Lubricant Additives

東工大（学）*野間 央 東工大（正）青木 才子 出光興産（正）小林 兼士

Hiroshi Noma*, Saiko Aoki*, Kenji Kobayashi**

*Tokyo Institute of Technology, **Idemitsu Kosan Co., Ltd.

1. はじめに

現在、様々な研究分野で機械学習の利用が進んでいる。トライボロジー分野においても、機械学習を利用したトルク信号による潤滑状態の予測¹⁾や摩擦表面画像から摩擦係数を予測する研究²⁾などが行われている。筆者らの研究グループでは、無段変速機油に着目し、リン、硫黄、カルシウム系の各種添加剤併用系におけるトライボフィルムの摩擦発現メカニズムについて調査を行ってきた。これまでの研究結果により、リン系添加剤にカルシウム系添加剤を併用することによりリン系添加剤単体より高摩擦を発現することを確認した。これはカルシウム系添加剤の併用によりリン系トライボフィルムの形成が促進して厚い被膜が形成するだけでなく、トライボフィルムを含む摩擦面の形状（モルフォロジー）がしゅう動方向と平行で平坦な筋目状の構造へと変化することが高摩擦の原因であると考察した。カルシウム系添加剤の併用が摩擦増大に起因することは明確であるが、トライボフィルムの元素分布や表面形状など形成状態と摩擦係数との関連性は明らかになっておらず、摩擦係数と関係性を持つ因子の統括的な検討は行われていない。そこで、摩擦係数の予測ができるだけでなく、予測することにより摩擦係数との関連性を明らかにすることができる機械学習を応用することを着想した。本研究では、複数の添加剤併用系で形成されたトライボフィルムの摩擦発現メカニズムを調査するために、摩擦試験により得られた実験データを機械学習に使用し、摩擦係数の予測と同時に摩擦に関与する因子の関係性を評価した。

2. 摩擦係数を予測する機械学習モデルの構築プロセス

2.1 データセットの準備とデータの事前処理

本研究では、各種試料油の摩擦試験により得られる摩擦係数と試験片摩耗痕に形成されるトライボフィルムの形状観察および元素分析結果を機械学習に使用した。まず、摩擦試験では、試料油として、基油には 100 °C における動粘度が 3.88 mm²/s であるポリ α オレフィン (PAO)、リン系添加剤としてリン酸トリクレジル (TCP, Tricresyl Phosphate)、硫黄系添加剤としてジベンジルジスルフィド (DBDS, Dibenzyl Disulfide)、カルシウム系清浄剤として過塩基性カルシウムスルホネート (OBCS, Overbased Calcium Sulfonate) を用いた。高速往復動摩擦試験機を用いて、試験温度 100°C、試験荷重 29 N として、ボール試験片 (SUJ2, ϕ 9.525 mm) を振動数 15 Hz、振幅 2.8 mm で往復しゅう動させ、ディスク試験片 (SUJ2, ϕ 8×8 mm) の間で生じる摩擦力を測定し摩擦係数を算出した。試験時間は 4 h とした。

次に、トライボフィルムの形状観察および元素分析では、ディスク試験片摩耗痕に形成したトライボフィルムを対象として、分析位置はトライボフィルムの中央部とした。共焦点レーザ顕微鏡によりトライボフィルムの形状観察を行い、各種粗さパラメータを測定した。また、電子線マイクロアナライザ (EPMA, Electron Probe Micro Analyzer) によりトライボフィルムの元素分析を行い、P と Ca のマッピング画像における元素強度分布からヒストグラムを作成し、一番多く出現する強度値である最頻値を算出した。

摩擦試験結果を踏まえ、機械学習により摩擦係数の予測を実施するに当たり、入力値には添加剤の濃度と摩擦試験により形成したトライボフィルムの各種粗さパラメータ、P と Ca の元素強度の最頻値を使用し、出力値には摩擦試験終了時の摩擦係数を用いた。データセットのサンプルサイズ n は 36 であり、データの偏りをなくすためにデータの事前処理として入力値を標準化させた。

2.2 機械学習モデルの構築

機械学習モデルを構築する工程として、データセットを訓練データとテストデータに分け、訓練データにより構築したモデルを用いてテストデータにより算出した予測値と実測値を比較することで、モデルの性能評価を行った。これは、すべてのデータセットを用いて機械学習モデルを構築した場合、そのデータセットに適した機械学習モデルが構築されるため、未知のデータでは性能の悪いモデルができる可能性があるためである。本研究において機械学習モデルを構築する際、データセットは 4:1 の割合で訓練データとテストデータに分類した。

次に、前処理を行ったデータセットから、機械学習を構築した。機械学習を構築するために用いられるアルゴリズムは様々あるが、本研究では、決定木 (DR)、アンサンブル学習であるバギング (BR)、ランダムフォレスト (RFR)、アダブースト (ABR) の 4 種類のアルゴリズムを使用して機械学習モデルを構築した。決定木はデータセットを分割し、枝分れのある樹形図を描くが、データを分割する際に、二つのクラスターの不純度が最も小さくなるように任意の特徴量の値を境界とする。また、決定木を用いたアンサンブル学習とは、決定木をいくつか作成し、それぞれの決定木で求めた値を平均させて予測値を算出する方法である。しかし、同じデータセットで決定木を作成しても同じ決定木しかできないので、複数選択できるデータセットを用いる方法や特徴量の数を制限する方法により弱い学習器を多

く作成することで、モデル性能を向上させることができる。今回のモデル構築に用いたプログラミング言語は Python であり、IDE の一種である Jupyter Notebook 上で実行した。

3. 結果および考察

3.1 機械学習モデルの評価

摩擦係数を予測するモデルを構築し、モデル性能の評価を行った。性能評価は、それぞれの学習器で算出される決定係数 R^2 で行い、それぞれの学習器で算出した R^2 を Fig. 1 に示す。決定係数の値が 1 に近似するほど、構築したモデルより算出された予測値と実測値が近似し、予測性能が高いことを示す。Figure 1 より、決定木や決定木を用いたアンサンブル学習であるランダムフォレスト、バギング、アダブーストは 1 に近い決定係数を示し、予測性能は高いことがわかる。このことから、決定木系の機械学習モデルは摩擦係数を予測に適していることが伺える。また、決定木系の学習器の特徴として、入力値同士が関係性を持つ場合に高い予測性能を示すことが知られている。本研究の結果より、決定木系は高い予測性能を示したことから、入力値同士が複雑に関係性を持っていることが推察される。

3.2 ランダムフォレストの特徴量重要度

ランダムフォレストの結果から算出される特徴量重要度の結果を Fig. 2 に示す。任意の特徴量で分岐する前後の不純度について、枝分れした前後のクラスターのサンプルサイズに全体のサンプルサイズを除した値をそれぞれ重み付けし、分岐する前後でその値が減少した分を特徴量重要度として算出した。すなわち、特徴量重要度は特徴量による精度向上の寄与率を知ることができるため、本研究のランダムフォレストの結果にも応用した。Figure 2 より、ランダムフォレストの特徴量重要度の結果において、Ca 強度、OBSC 濃度が顕著に高い値を示し、これらの特徴量が摩擦係数を予測するために重要な特徴量であることがわかった。

3.3 ネットワーク分析結果

本研究で用いたデータセットから算出されたネットワーク分析結果を Fig. 3 に示す。ネットワーク分析では、データセットのそれぞれの特徴量における相関係数が算出されるため、相関係数が高い特徴量同士を紐づけて関連性を可視化することができる。すなわち、ネットワーク分析により、摩擦係数と関係性の高い特徴量だけではなく、他の特徴量同士の関係性を見ることができると推測される。Figure 3 より、摩擦係数との相関係数が高いものは、ランダムフォレストの特徴量重要度の結果において摩擦係数予測に重要であると示唆された Ca 強度、OBSC 濃度が含まれていることがわかった。また、相関係数が高い特徴量は複雑に絡み合っていることも明らかになり、これが本研究のデータセットにおいて決定木系が摩擦係数の高い予測性能を示した原因であることが考えられる。

4. まとめ

- 1) 決定木系の機械学習モデルは摩擦係数の高い予測性能を示した。
- 2) 特徴量重要度の結果より、Ca 強度、OBSC 濃度は摩擦係数を予測するための重要な特徴量であることが示された。
- 3) ネットワーク分析結果より、特徴量は複雑に絡み合い、これが決定木系による摩擦係数の高い予測性能を示す原因と考えられる。

文献

- 1) Jakob. M, et al. 2018, *Lubricants*, 6(4), 108
- 2) 山田高澄・他著者ら, 2021, トライボロジー会議東京予稿集, 378

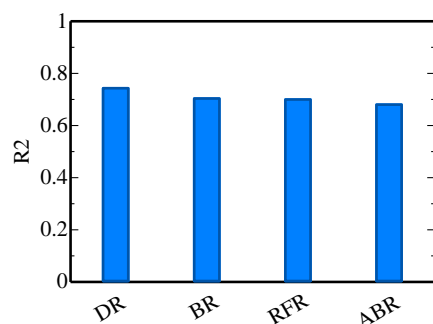


Fig. 1 Performance results of models

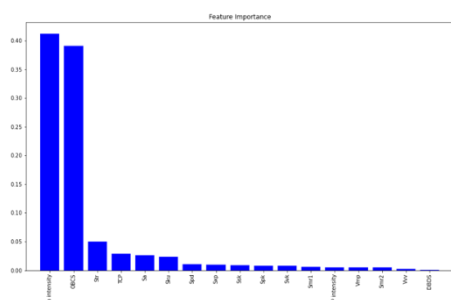


Fig. 2 Feature importance by random forest

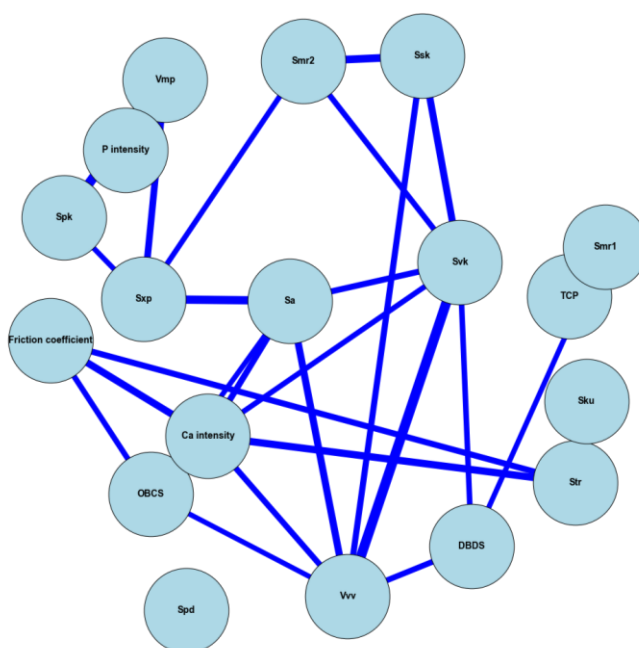


Fig. 3 Network result of the dataset