

潤滑油の組成と性能を関連付ける数理モデルの考察

Exploring Mathematical Models Linking Lubricant Formulation and Performance

出光興産(正) *小林 兼士

Kenji Kobayashi

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

1. 目的

潤滑油の組成を説明変数、性能を目的変数として適切な数理モデルを組み立てることは、統計解析により組成と性能の関係を把握したり、機械学習にて性能を予測したりすることに役立てられる。本報では、潤滑油の原材料配合量を説明変数、耐摩耗性評価の結果を目的変数として、数値を予測する回帰問題と、グループを予測する分類問題の2つの課題を設定した。本報の目的は各課題にて潤滑油の組成と性能を関連付けるのに適した数理モデルについて考察することで、統計解析や機械学習に役立つ知見を得ることである。

2. 回帰問題を題材にした数理モデルの比較¹⁾

潤滑油はベースオイルと添加剤から構成される。本検証で用いた試料油のベースオイルは2種類のどちらか1つを選択し、添加剤は4種類から選択し配合した。ベースオイルは、低粘度鉱油(以下、LV-BO, 100℃動粘度 2 mm²/s)と高粘度鉱油(以下、HV-BO, 100℃動粘度 6 mm²/s)から選定した。添加剤はリン酸トリクレジル(TCP, Tricresyl Phosphate)、ジベンジルジスルフィド(DBDS, Dibenzyl Disulfide)、過塩基性カルシウムスルホネート(OBCS, Overbased Calcium Sulfonate)、飽和脂肪酸(SFA, Saturated Fatty Acid)の4種類を選定した。各添加剤の配合量はそれぞれ0.00wt%, 0.25wt%, 0.50wt%の3水準とした。2種類のベースオイルから1つ選び、4種類の添加剤を3水準の濃度で独立して選定した場合の配合量の組み合わせは162通りであるが、その組み合わせから選定する原材料・濃度の数が可能な限り等しくなるよう50通りの配合量(内3通りは重複)を抽出した。選定した50通りの配合量の組み合わせをランダムに並び、耐摩耗性を順次評価した。耐摩耗性は四球摩耗試験(ASTM D4172 参考)により、各試料油を同条件で評価した。試験球摩耗痕径が小さいほど、耐摩耗性が良好であると判断した。各原材料の配合量を説明変数(x)としたが、ベースオイルはLV-BOを0, HV-BOを1とするダミー変数を説明変数とした。摩耗痕径を目的変数(y)として重回帰モデルと決定木モデルの2つを作成した。

重回帰モデルは式(1)で示す重回帰式にて説明変数の各係数(a)を最適化した。決定木モデルは、データセットを2つに分割する操作を繰り返して、説明変数と目的変数の関係を表現したモデルであるが、2つに分割するにあたっては、分割前後の平方残差和(RSS, Residual Sum of Squares)の差が最大になる基準にて分割を行った。この決定木モデルは過学習を防ぐため、決定木のノード分割数をパラメータとして情報量規準 AIC (An Information Criterion)を算出し、AICが最小となる決定木のノード分割数で得られたモデルを採用した。

重回帰モデルの結果として、回帰式による予測値と実測値の関係を Fig.1 に示す。決定係数 R^2 は 0.16 と小さく、本検証において重回帰モデルは潤滑油の組成と性能の関係を適切に表現できないと判断した。決定木モデルの結果を Fig.2 に示す。ノード分割数は AIC が最小となる 4 とした。決定係数 R^2 は 0.35 であり弱い相関であるが、重回帰モデルより良好であった。また Fig.2 に示す決定木モデルから、低粘度ベースオイルを選択した場合において、SFA と TCP を組み合わせると一定の配合量以上配合することで良好な耐摩耗性が得られるという仮説を立てることができ、その仮説が妥当であることを実測により確認した¹⁾。

$$y = a_0 + a_A x_A + a_B x_B + \dots + a_n x_n \quad \dots (1)$$

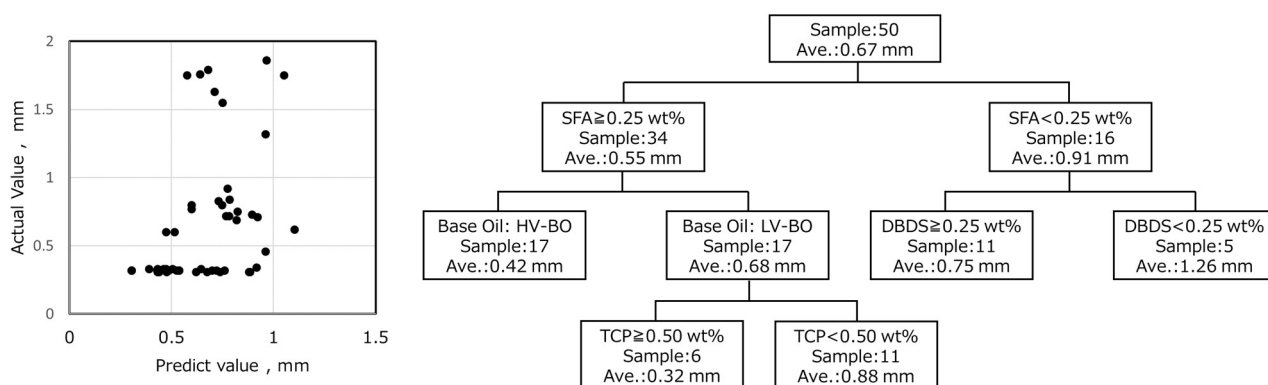


Fig.2 Result of decision tree analysis

3. 分類問題を題材にした数理モデルの比較²⁾

前項で述べた回帰問題とは異なるデーセットにより試験結果のグループを予測する分類問題を設定した。検証には 741 個の潤滑油の原材料配合情報と四球摩耗試験(ASTM D4172 参考)の試験結果(摩耗痕幅, mm)を用いた。試験結果は 0.50mm を基準として二値化し, 0.50mm を超えるものを不良(ダミー変数「0」), 以下のものを良好(ダミー変数「1」)とし 2 つのグループに分けた。741 個の試験結果から, 20 個の試験結果を予め抽出し, 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。学習用データから各数理モデルで予測モデルを構築した。そして構築した予測モデルによる検証用データの予測結果を実測結果と照合し, その正答率を数理モデル毎に比較した。数理モデルとしてロジスティック回帰モデル, ニューラルネットワーク(NN, Neural Network) モデル, ランダムフォレスト(RF, Random Forest) モデルの 3 つを選択し予測精度を比較した。

ロジスティック回帰モデルによる予測を行うにあたり, 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回帰式の係数を最適化した。NN モデルには様々なパラメータがあるが, 今回は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。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, 十分な予測精度が得られるよう試行錯誤により決定した。RF モデルは, ランダムに複数の決定木モデルを生成し, 各決定木の出力結果の多数決により予測結果を出力するモデルである。決定木を分割するための指標としてジニ不純度を用いた。1 つの決定木モデルに用いる決定木のノード数や, 決定木の分割で用いる特徴量の数などのハイパーパラメータは十分な予測精度が得られるよう試行錯誤により決定した。NN モデルおよび RF モデルは予測精度が乱数に依存し試行毎に正答率が異なるため, 3 つの乱数設定により正答率を求めて予測精度を評価した。

各手法における正答率を数理モデルのイメージと共に Fig.3 に示す。ロジスティック回帰を用いた場合の正答率は 50%であり, 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ることから十分な予測精度とは言えなかった。一方で NN 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, 今回検証した 3 つのモデルの中で最も良好な予測精度であった。RF モデルの正答率は 60%から 70%であり NN モデルに劣る結果であったが, 各原材料の特徴重要度を計算し, 各原材料が試験結果に与える影響を定量化できた。

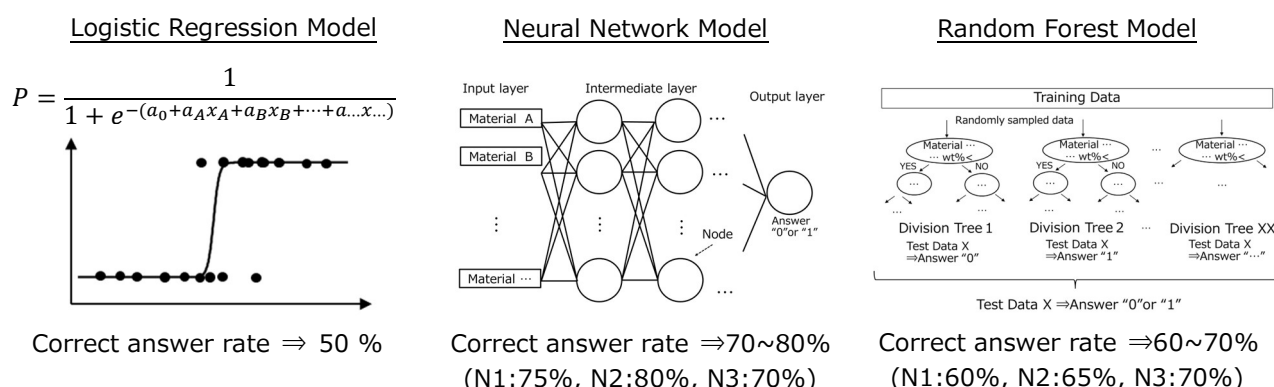


Fig.3 Image of three mathematical models and correct answer rate

4. まとめ

全ての原材料が一相に混合されることで全ての原材料が性能に関与し得る形態の潤滑油において, 原材料同士の複雑な交互作用が性能に影響を及ぼすことが想定される。そのため潤滑油の組成と性能の関係は各原材料の影響を足し合わせるような線形モデルではなく, 2 つ以上の原材料による交互作用も含めることができる非線形モデルで表現することが適切であると考えられる。今回設定した 2 つの課題の中で, 重回帰モデルやロジスティック回帰モデルと比べて決定木モデル, NN モデル, RF モデルが比較的良好な予測精度を与えた。その理由としてこれら 3 つのモデルが潤滑油の組成と性能の非線形な関係を学習できたことが挙げられる。潤滑油の組成と性能を関連付ける数理モデルについてさらに理解を深めるにあたっては, 耐摩耗性以外の性能に対する数理モデル構築や, 元素量や化学構造といった組成物の配合量以外のパラメータを説明変数とした数理モデル構築などが, 今後取り組むべき課題と考える。

- 1) 小林：潤滑油の耐摩耗性能予測における機械学習の適用 第 2 報, トライボロジー会議 2024 春 東京 予稿集, (2024) B32
- 2) 小林：潤滑油の耐摩耗性能予測における機械学習の適用, トライボロジー会議 2023 春 東京 予稿集, (2023) D30