

モデルベース開発に向けたデータマイニング技術 による車両性能設計システムの研究

Vehicle Performance Design System Using Data Mining Techniques
towards Model-based Development

有 吉 斗紀知*
Tokitomo ARIYOSHI
福 重 貴 浩*
Takahiro FUKUSHIGE

春 松 正 敏*
Masatoshi HARUMATSU
山 田 健太郎*
Kentaro YAMADA
矢 野 和 重*
Kazushige YANO
鳴 川 要**
Kaname NARUKAWA

要 旨

ABSTRACT

車両開発初期段階で各性能の目標値割り当てが可能な車両性能設計システムを構築した。本システムは、モデルベース開発の考えに基づき、適切な目標値の決定をデータマイニング技術の適用により実現している。

本システムは車両開発時のデータを用いて、以下三つの技術から構成される。

- (1) 少数データをもとに高精度な予測を実現する応答曲面生成技術
- (2) ユーザの意思を反映した高効率なデータ探索ができる対話型データ選好技術
- (3) 性能と諸元の関係から、性能に大きく影響する諸元がわかる寄与度分析技術

構築したシステムを車両運動性能シミュレーションに適用することで、簡単かつ高速に目標とする性能値や諸元値を決定できることを確認した。

A vehicle performance design system has been developed that makes it possible to assign target values for different performance demands at the initial stage of the vehicle development process. Based on the concept of model-based development, this system determines appropriate target values through the application of data mining techniques.

The system uses the data available at the time of vehicle development, and comprises the following three techniques:

- (1) A response surface generation technique that helps enable high-accuracy predictions based on small amounts of data;
- (2) A decision-making technique that helps enable high-efficiency data searches that reflect the user's intentions; and
- (3) An impact analysis technique that makes it possible to identify specifications that have the great effect on performance from the relationships between performance and specifications.

It was demonstrated that the application of this system to vehicle dynamic performance simulations enables determination of target performance values and specification values simply and at high speed.

1. まえがき

今日、自動車設計には環境性能や安全性能などへの配慮から様々な要求が課されるようになっており、目標性能実現のための各要素の開発は大規模、複雑化している。一方で、開発期間の短縮とコスト削減は引き続き主要な課題となっている。

また、自動車市場は世界的に見ると、年率 4 % で拡大し

* 四輪 R&D センター

** 基礎技術研究センター

ており、新興国などの市場拡大等による地域市場の販売構成変化が生じている。この状況に対応するために、ラインナップの増加が必要であり、自動車業界全体で、開発フローの改善が行われている。とりわけデジタル開発が活発化してきており、実機試作をできるだけ行わず、シミュレーションや実験データ等を用いたバーチャル開発を行うモデルベース開発 (Model Based Development: MBD) の重要性は高まっている。

本研究では、MBDを実現する車両性能設計システムを構築することにより、開発初期段階において、燃費、加速性能など複数の車両目標性能と、それに紐づく諸元値を導き出すことで、開発効率の向上を目指す。

2. システム構成

本研究では、データマイニング技術を用いて、次世代自動車開発に貢献する車両性能設計システムを構築し、複数性能のトレードオフを考慮しながら適切な目標値を設定するという開発スタイルを実現する。

開発したシステム（以下、本システム）はデータベース（Data Base: DB）を有するデータドリブン型シミュレーション^{(1), (2)}であり、精度と応答性が重要となる。Fig. 1に車両性能設計システムの概要を示す。

Fig. 1に示すように、本システムは、車両性能を計算するコンピュータ支援エンジニアリング（Computer Aided Engineering: CAE）計算部、DBを含む目標性能表示部と三つのデータマイニング技術から構成される。その実行手順を以下に示す。

- (1) 本システムの入力となる諸元値を実験計画法でサンプリングする。
- (2) CAE計算を実施することにより、教師データを作成する。
- (3) 作成した教師データを利用し、機械学習の技術を用いて応答曲面とDBを構築する。
- (4) 構築したDBを利用して複数性能のトレードオフを考慮しながら適切な目標値を決定する。

このDBはCAEの入力として使用された諸元値と計算結果の性能値で構成される。左5列の諸元値と右3列の性能値からなるDB配列の例をTable 1に示す。このDBを利用することにより車両性能設計システムを実行することができる。以下(1)-(3)に車両性能設計システムを実現するデータマイニング技術の概要を示す。

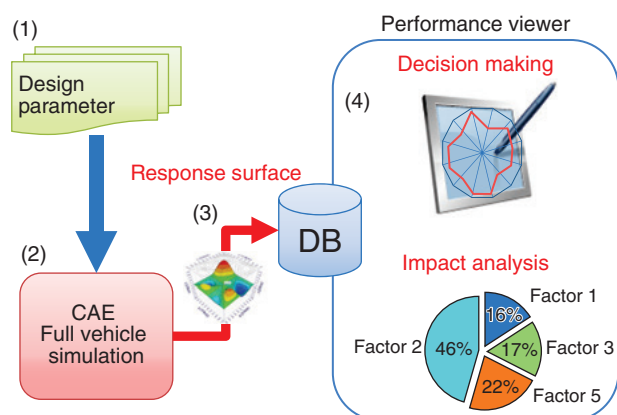


Fig. 1 Vehicle performance design system

Table 1 Database matrix

	Factor					Performance		
	1	2	3	4	5	1	2	3
1	460.91	2026.05	0.42	1418.89	2533.81	6.15	136	4.34
2	448.78	1775.30	0.42	1368.02	2511.51	6.19	137	4.41
3	459.69	2204.39	0.44	1417.41	2500.43	4.81	135	4.26
4	439.90	1924.32	0.44	1346.49	2487.21	5.13	137	4.37
5	458.57	1850.84	0.42	1348.45	2527.38	5.07	136	4.35
...	...	...	...	...	...	...	...	...
200	450.60	1745.68	0.44	1409.86	2590.14	5.38	138	4.39

(1) 応答曲面生成 (Response surface)

本システムは、目標性能設計を行うためにCAE計算結果の大規模なデータを有するDBを構築する必要がある。

10万のケースデータを作成し、DBを構築するための従来手法と応答曲面近似方法の比較をFig. 2に示す。Fig. 2は従来手法と提案手法の計算の違いを示しており、従来通りのCAE計算でDBを構築するためには非現実的な計算時間を要することがわかる。そのため、できるだけ短時間でDBを構築する技術が必要である。提案手法では200ケース程度の少ないCAE計算結果を教師データとし、機械学習の技術を用いて応答曲面 (Response Surface Model: RSM) を生成する。その応答曲面を用いて、10万ケース分のデータを計算することによって、従来に比べ計算時間を大幅に短縮できる。詳細は2.1.節で述べる

(2) 対話型データ選好 (Decision making)

構築した10万ケースのDBから、ユーザの意図を反映した適切な解を導出することは容易ではない。そこで本システムは、以下の二つのデータ選好手法を構築した。

(a) 性能値からダイレクトに諸元値を探索するダイレクト探索手法

(b) クラスタリング技術を用いてデータの傾向を見ながら対話的に絞り込みできるグループ探索手法

上記二種類の探索手法からユーザは用途に合った対話型データ選好手法を選択して実行することが可能である。詳細は2.2.節で述べる。

(3) 寄与度分析 (Impact analysis)

寄与度分析は、データの線形・非線形性や複数の諸元値を有するデータ環境でも、性能と諸元値の特性関係を把握できるようにする分析手法である。

変数間の非線形関係が複数諸元にわたる場合、関係性を可視化するのは容易ではない。本システムでは複数諸元で

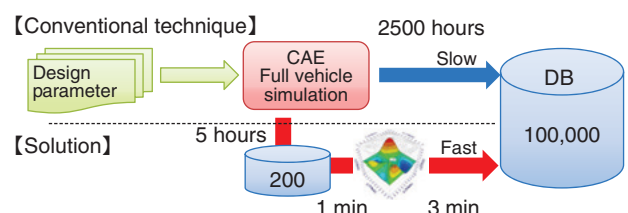


Fig. 2 Effect on response surface

も特性の表示ができる部分従属プロット⁽³⁾を使って可視化している。また各諸元値が性能に影響する度合いを、直観的に把握しやすい寄与度として算出しグラフで表示するようにした。詳細は2.3.節で述べる。

2.1. 応答曲面生成

応答曲面は、入力データとしての諸元値 $\mathbf{x}$ に対しシステムの出力データとしての性能値 $\mathbf{y}$ を対応させる数学モデルである。システムの例としては、実験系あるいは工学シミュレーションなどがある。往々にして、実験やシミュレーションを実行するには時間的、人的コストを要するため、ケース数をあまり多く得ることができない。そこで、少ないケースデータをもとに応答曲面を生成することができれば、新たな入力 $\mathbf{x}$ に対する出力 $\mathbf{y}$ を簡単に予測することが可能である。なお応答曲面の生成に必要なケースデータは、教師データと呼ばれる。

モデル化の誤差を ε とおけば、入出力間の関係は式(1)であらわされる。

$$\mathbf{y} = M(\mathbf{x}) + \varepsilon \quad (1)$$

ただし、 $M(\mathbf{x})$ は応答曲面である。一般に出力 $\mathbf{y}$ はベクトル量だが、各成分に対しモデル化が可能であるため、以下では出力 $\mathbf{y}$ はスカラー量とする。

応答曲面 $M(\mathbf{x})$ の生成手法は様々あるが、本研究では、少ない教師データから近似精度の高いモデルを生成する必要があることから、PCE-LAR 法⁽⁴⁾と Kriging 法の組合せである PCK 法を採用した。PCE-LAR 法で生成された関数を $f(\mathbf{x})$ 、Kriging 法で生成された関数を $Z(\mathbf{x})$ とおけば、 $M(\mathbf{x})$ は式(2)であらわされる。

$$M(\mathbf{x}) = f(\mathbf{x}) + Z(\mathbf{x}) \quad (2)$$

Fig. 3 に示すように、PCE-LAR 法は大局的な出力の変化をあらわし、Kriging 法は局所的な出力の変化をあらわしている。大局と局所の両方のモデル化により高い近似精度を実現できる⁽⁵⁾。以下、PCE-LAR 法および Kriging 法の詳細を説明する。

2.1.1. PCE-LAR 法

PCE-LAR 法は、多項式カオス展開法 (Polynomial Chaos Expansion 法: PCE) による大局的なシステム応答の

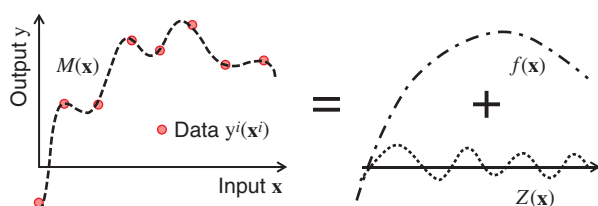


Fig. 3 Response surface model $M(\mathbf{x})$ as sum of global trend $f(\mathbf{x})$ and local fluctuation $Z(\mathbf{x})$

モデル化と最小角回帰法 (Least Angle Regression: LAR) による近似の高精度化を組み合わせた手法である。PCE 法では、モデル $f(\mathbf{x})$ を直交多項式の線形結合としてあらわす。スカラー入力 x の場合、 $f(x)$ は式(3)で与えられる。

$$f(x) = \sum_{i=0}^N a_i \Phi_i(x) \quad (3)$$

ただし、 a_i は係数、 N は多項式の最高次数、 $\Phi_i(x)$ は i 次の正規直交多項式である。本研究では、直交多項式にルジャンドル多項式を使用した。ルジャンドル多項式の入力 x の変域は $[-1, 1]$ であるため、当該変域に入力値をあらかじめ変換しておく必要がある。式(3)においては、1 から N までの全てのルジャンドル多項式が使用される。

係数 a_i は一般化最小二乗法により決定する。教師データ点 $(\mathbf{x}, y)$ の点数を n とすれば、 n は係数の個数以下でなければならない。ところが、入力 $\mathbf{x}$ の成分数 M あるいは最高次数 N の増加により、係数の個数は爆発的に増加する。例えば $M = 10, N = 4$ の場合、4 次までのルジャンドル多項式すべてを考慮すると、モデルに必要な係数は 1000 個を超える。これは、 M あるいは N の増加により、必要な教師データ点数が爆発的に増えることを意味する。

この課題を回避するのが LAR アルゴリズムである。LAR アルゴリズムは、教師データを参照しながら、モデルを構成するのに適切な次数の多項式を、優先順位をつけて判別する。これによりモデルに使用する $\Phi_i(\mathbf{x})$ の個数、つまり係数の個数を大幅に削減することができる。また、この削減によりモデルの近似精度も向上する。その理由は、LAR アルゴリズムを適用しない場合、項数の爆発的な増加のために高次多項式まで使用することが容易ではないのに対し、LAR アルゴリズムを使用すると、必要な次数の多項式のみを使用してモデルを構成するため、項数爆発を招くことなく高次多項式を使用できるからである。

2.1.2. Kriging 法

Kriging 法は平均からの変動量を記述する。この変動は確率過程に従っているとみなし、異なる二点の教師データ $\mathbf{x}^i$ と $\mathbf{x}^j$ における変動は、距離 $|\mathbf{x}^i - \mathbf{x}^j|$ に依存する相関関数 R_{ij} をもつとする。 R_{ij} には、Gauss 型、指数型、Matérn 型があるが、本研究では教師データに応じて適切な型を判別するようにした。どの型にもパラメータが一つ含まれるため、最適化計算によりその値を決定する。

Kriging 法によりモデル化された変動 $Z(\mathbf{x})$ は式(4)で与えられる。

$$Z(\mathbf{x}) = \mathbf{r}(\mathbf{x}) \mathbf{R}^{-1} \Delta \mathbf{y} \quad (4)$$

ここで、 $\mathbf{r}(\mathbf{x})$ は $\mathbf{x}$ と教師データ $\mathbf{x}^i$ との相関で構成される n 成分の行ベクトル、 $\mathbf{R}$ は R_{ij} を成分とする $n \times n$ 行列、 $\Delta \mathbf{y}$ は教師データ y^i から $f(\mathbf{x}^i)$ を引いた n 成分の列ベクトルである。

実際の計算では、応答曲面の構築は前段と後段に分けて実行する。前段では、LAR アルゴリズムにより $f(\mathbf{x})$ で使用する多項式 $\Phi_i(\mathbf{x})$ の候補順位を決定する。この時点ではまだ $f(\mathbf{x})$ を決定しない。理由は、PCE 法単体で最適化した場合の $f(\mathbf{x})$ と、Kriging 法との組み合わせで最適化した場合の $f(\mathbf{x})$ とは、必ずしも一致しないからである。後段では適切な相関関数の型をまず判別する。次に前段で決めた優先順位に従い、 $f(\mathbf{x})$ で使用する $\Phi_i(\mathbf{x})$ を一つずつ追加しながら Kriging 法との組み合わせで決まる $M(\mathbf{x})$ の誤差評価を行う。最小の誤差を与える $f(\mathbf{x})$ と $Z(\mathbf{x})$ の組み合わせを用いて、応答曲面 $M(\mathbf{x})$ を決定する。

2.1.3. 手法の検証

開発した手法を燃費および加速シミュレーションから得られたデータに適用し、近似誤差を検証した。燃費シミュレーションから得られたモード燃費のうち、三種類 (US-Hot, JC08, EUDC) に対しモデル化を行った結果を Fig. 4 に示す。参考として、市販の応答曲面生成ツールで精度の良かった手法 (異方性 Kriging 法: AKR) の結果も示した。図の横軸は応答曲面の生成に使用した教師データの点数、縦軸は相対誤差 (%) である。相対誤差の算出にあたっては、教師データとは異なるデータを使用し、各データ点における相対誤差を求め全点にわたり平均した。いずれの燃費モードにおいても、AKR $\rightarrow$ PCE-LAR $\rightarrow$ PCK の順に近似誤差が小さくなっている。加速シミュレーションデータへの適用については、二種類の加速モード 0-100 km/h 加速および 80-120 km/h 加速での加速時間に対してモデル化した結果を Fig. 5 に示す。誤差の大きさは燃費データの場合に比べ半分以下となり、手法間の違いでは、燃費の場合と

同様に AKR $\rightarrow$ PCE-LAR $\rightarrow$ PCK の順に誤差が小さくなる結果となった。

上記の結果では、燃費および加速シミュレーションにおいて、PCE-LAR 法に比べ PCK 法で誤差の改善が認められた。両手法の差を理解するため、Fig. 6 にモード燃費シミュレーションから得られたモデルを示す。横軸は正規化した設計値、縦軸は基準値まわりの変動を表現したモード燃費値である。PCE-LAR 法は、入力変数の全領域に対して全体的な傾向をうまく捉えることができている。しかし、局所的にはデータとモデル予測値が大きくかい離している場合がある。PCK 法では、Kriging 法との併用でこのかい離が大きく改善できていることがわかる。データの特徴によっては、これほど明らかな改善が見られないこともあるが、PCE-LAR 法単体あるいは Kriging 法単体で生成した応答曲面よりも、両者を併せた PCK 法の方がロバストな性能を示す傾向にある。

2.2. 対話型データ選好

大規模 DB を利用して、車両性能目標値を決定したいが、適切な車両性能データを容易に見つけられないことが課題である。そのために、ユーザは対話型データ選好を行い、高速かつ少ない操作回数で車両性能データの探索を行う。

本研究では、探索方法にダイレクト探索とグループ探索の二つの手法を採用した。

2.2.1. ダイレクト探索 - 高速探索法

高速探索法とは、DB 内の車両性能データと車両性能目標値のユークリッド距離を計算し、距離の短いデータを解とする探索方法である。Fig. 7 に 3 次元データを例とした探索手法を示す。ユーザは★のデータを目標とするが、データ空間が疎であるために、最も近似しているデータを探索解としている。このように、構築された DB 空間にはデータの疎密があるため、適切な探索解が DB 内になれば、解の精度は低くなるが、近似探索であるために高速に計算が可能である。

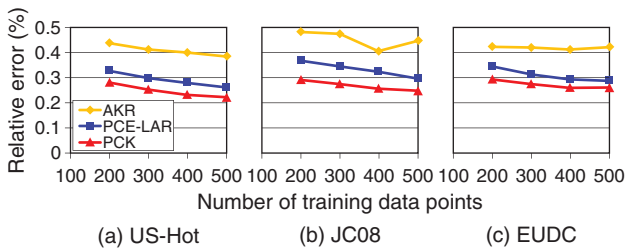


Fig. 4 Relative error plot for three fuel economy modes

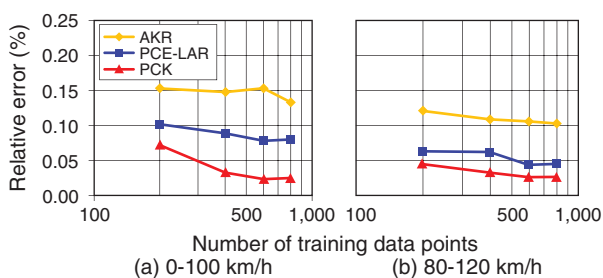


Fig. 5 Relative error plot for two acceleration modes

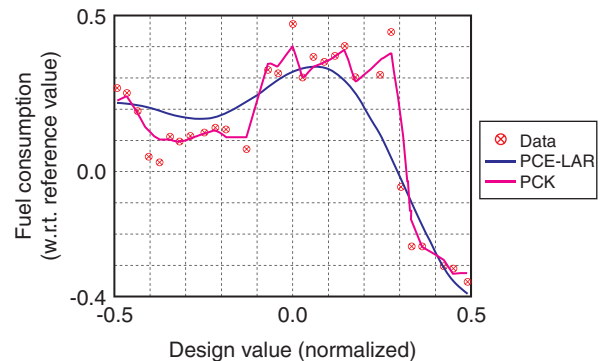


Fig. 6 Comparison of simulation data and response surface model predictions (PCE-LAR and PCK)

2.2.2. ダイレクト探索 - 高精度探索法

高精度探索法とは、大規模 DB を構築するために作成した応答曲面を利用した最適解探索を行う方法である。Fig. 8 に 3 次元データから応答曲面を構築し、解の探索に利用した探索手法を示す。ユーザは★のデータを目標としており、本手法は DB 空間内のデータ間を数学的に補間しているため、その補間情報を基に最適な値を計算することで探索解を求めている。なお、勾配法を用いて応答曲面計算結果の性能値と車両性能目標値の誤差が 0 に近似するデータを探索している。

課題として、初期値の影響により局所最適化が行われることが考えられるために、真の最適解となれない場合が生じるが、あらかじめランダムに初期値を 100 回変化させて局所最適化の回避を行い、真の最適化が行われるようにしている。この手法は、最適化問題を解くために計算時間は長くなるが、解の精度は高いといえる。

2.2.3. グループ探索

グループ探索とは、DB 内のデータをクラスタリングし

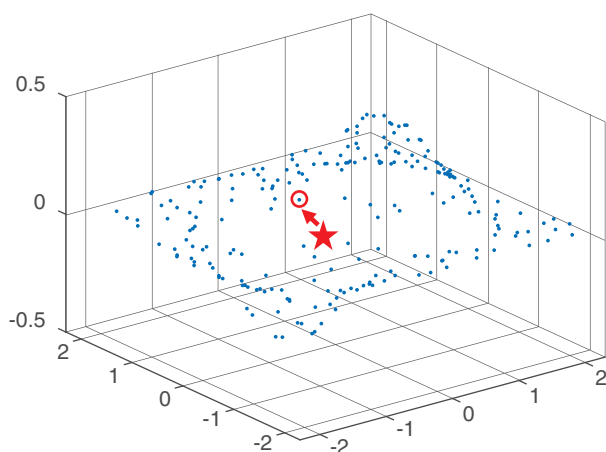


Fig. 7 High-speed search method by data exploration

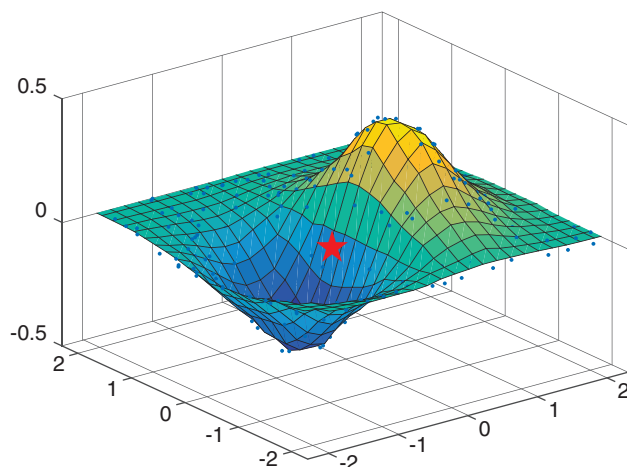


Fig. 8 High-accuracy search method by response surface

てクラスの重心データである代表データをレーダチャートに表示する。ユーザはレーダチャートを利用して、各データを見ながら絞り込みを行うことで解の探索が可能である。データ絞り込み例を Fig. 9 に示す。DB 内のデータをクラスタリングすると、Fig. 9 の左側に示すように五つのクラスで表現され、ユーザはその中から一つのクラスを選択する。その選択結果を Fig. 9 の右側に示す。各レーダチャートの項目には、DB が有する性能値 (P1-P12) が示されているために、ユーザは全体の性能バランスを見ながら、所望の性能値の絞り込みを行うことが可能である。

本研究で扱うクラスタリング手法は、ユーザのデータ選択の幅を広げるため、Fuzzy C-Means⁽⁶⁾を導入した。

Fuzzy C-Means 手法は、従来のクラスタリングアルゴリズムの前提であるクラスの従属判定を行わず、あるデータが一つのクラスに完全に属するのではなく、他の複数のクラスにも少しずつ属していると考え、クラス所属の帰属度と呼ばれる度合いを用いて、分離境界面を曖昧に表現したものである。Fig. 10 に従来手法と Fuzzy C-Means 手法による三つのクラスの分離境界面の違いを示す。Fig. 10 の左側の図は一般的な従属判定を行ったクラスタリング結果であり、右側の図はクラスの分離境界周辺のデータに帰属度を使用した曖昧なクラスタリング結果である。

以下に Fuzzy C-Means のアルゴリズムを示す。

- (1) 要素 i のクラス j に対する帰属度 u_{ij} をランダムに初期化する。
- (2) 現在の帰属度 u_{ij} を用いて各クラスの重心ベクトル $\bar{x}_i$ を計算する。 m は、 $m > 1$ を満たすファジィさを決めるファジィパラメータ、 n は学習する特徴ベクトル数、要素 x_i はデータの集合である。重心ベクトル $\bar{x}_i$ は式 (5) で与えられる。

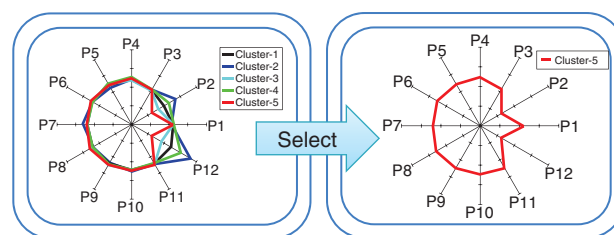


Fig. 9 Decision making

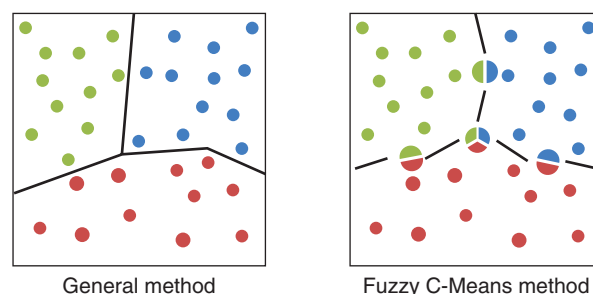


Fig. 10 Difference between clustering methods

$$\bar{x}_i = \frac{\sum_{j=1}^n (u_{ij})^m x_j}{\sum_{j=1}^n (u_{ij})^m} \quad (5)$$

- (3) 計算した重心ベクトル $\bar{x}_i$ を用いて要素 x_j のクラス i への帰属度 u_{ij} を更新する。 $\|\cdot\|$ はユークリッドノルム、 c はクラス数のパラメータである。帰属度 u_{ij} は式 (6) で与えられる。

$$u_{ij} = \left(\sum_i^c \left(\frac{\|x_j - \bar{x}_i\|}{\|x_j - \bar{x}_i\|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (6)$$

- (4) クラスごとに重心を再計算し、すべてのクラス重心が変化しなくなればクラスターリングを完了する。なお、変化すれば、(2) から再び実行する。

2.3. 寄与度分析

性能値は複数の諸元値の関数になることが多いため、各変数の影響を把握するのが容易ではない。そこで部分従属プロットを用いて性能と各変数との関係性を視覚化する。

部分従属プロットとは着目変数以外の変数の影響を平均化して表示するグラフである。Fig. 11 に部分従属プロットの例を示す。ここでは性能 y が五つの諸元で表される例を示しており、横軸に各諸元、縦軸に性能 y をとった五個の部分従属プロットで構成されている。

y は変数 1-5 の関数なので、たとえば変数 1 をある値に固定しても他の変数 2-5 を下限値から上限値まで変えると y の値は変動する。変動する y の値の平均値を求め、変数 1 の値に対する値としてプロットする操作を繰り返していくと黒実線で示す特性が得られる。これは変数 1 に対する y の平均的な変化特性を表す。さらに二本の赤点線は y の変動幅の 25 パーセンタイル位置と 75 パーセンタイル位置を示しており、他変数による y の変動の大きさの目安を表す。変数 2-5 についても同様な処理で部分従属プロットを作成できる。

Fig. 11 より、変数 2 に対する y の特性がほぼ直線であることから線形の関係があることがわかる。一方、変数 1, 3, 5 に対する y の特性が曲線になっていることからこれらの各変数に対して非線形の関係があることがわかる。さらに変数 4 に対しては特性がほぼ水平になっていることから性能にほとんど影響しない変数であることがわかる。二本の赤点線の間隔を見ると、どの変数も他の変数の影響を

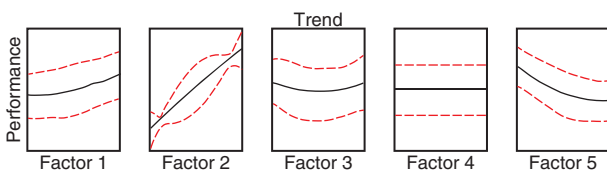


Fig. 11 Trend curves for parameters

受けることがわかる。変数 2 に対しては、赤点線の間隔が狭くなっている部分が二か所あり、ここで他変数の影響が小さくなっていることがわかる。

各諸元の影響の大きさを更にわかりやすく定量的に比較するには寄与度を用いる。寄与度は諸元を変化させたときの性能値の変化を表す指標であり、性能値を諸元で偏微分すると得られる。応答曲面関数を f とすると変数 x_i ($i=1, 2, \dots$) の寄与度 g_i ($i=1, 2, \dots$) は次のように表される。

$$g_i = \frac{\partial y}{\partial x_i} \cong \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_i} \quad (7)$$

寄与度も性能値と同様に多変数の関数になるため変数間で寄与度を比較する際に不便である。そこで g_i の平均値を変数 x_i の寄与度として用いることにする。

Fig. 12 は得られた寄与度を大きい順に棒グラフで表したものである。この例では変数 2 の寄与度が最も大きく、変数 1, 3, 5 の 2-3 倍程度あることや、変数 4 の寄与度がほぼ 0 であることがわかる。部分従属プロットとは異なり、各変数の影響の大きさや比率の比較に向いている。

さらに Fig. 13 に寄与度の割合を帯グラフで示す。変数 2 の寄与度の大きさが変数 5, 3, 1 を合わせた寄与度にほぼ匹敵することがわかる。

設計変更時の変数選択や数値変更量は、これらの寄与度情報に基づいて決定する。

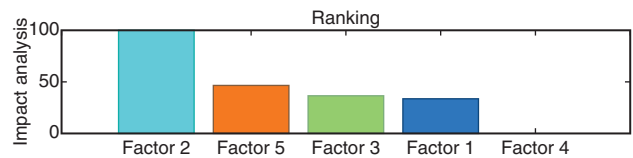


Fig. 12 Ranking of contribution ratio

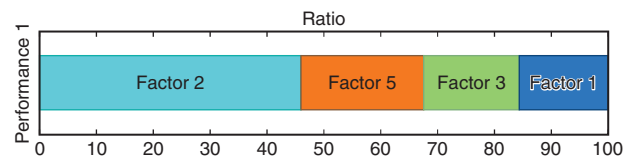


Fig. 13 Contribution ratio expressed as band graph

3. 適用

一例として、本研究にて開発した車両性能設計システムを利用した車両運動性能シミュレーション解析を示す。

はじめに、入力となる五つの諸元値データを実験計画法でサンプリングし、教師データ作成のために、CAE 計算で性能値を算出した。本適用例では 400 ケースのデータを用意した。ここで完成した性能値、諸元値のデータセットを用いて、大規模 DB 作成のために応答曲面を作成する。

応答曲面生成には2.1.節で示したPCE-LARを用いる。この例では三つの性能値に対し、三種類の応答曲面を生成する。作成した応答曲面を利用することにより、大規模DBを高速に計算できるだけでなく、性能から諸元を計算することができ、かつ、寄与度の分析ができる。以下に三つの解析結果を記す。

3.1. ダイレクト探索

三つの性能に対して、ユーザは目標とする値を指定する。一例として、三つの性能値に対する諸元値を計算する。目標とする三つの性能値からなるレーダチャート表現をFig. 14に示す。

Fig. 14で示した三つの目標値に対する諸元値計算結果をFig. 15, Fig. 16に示す。ここではダイレクト探索の高速探索法の結果をFig. 15に示し、高精度探索法の結果をFig. 16に示す。Fig. 15, Fig. 16ともに、五つの諸元値で構成されるレーダチャートで表現されており、Fig. 14で指定した目標値を使った五つの案別解を示している。

ユーザは単純に、目標とする性能値を指定するだけで、本システムは性能を満たす諸元値の組み合わせをユーザに提示する。案別解計算の所要時間は同一ワークステーション上で、高速探索法が0.043 sec, 高精度探索法が13.89 secであった。

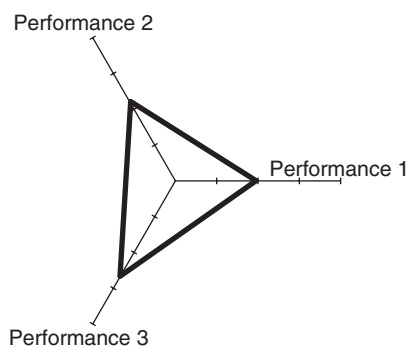


Fig. 14 Target performances

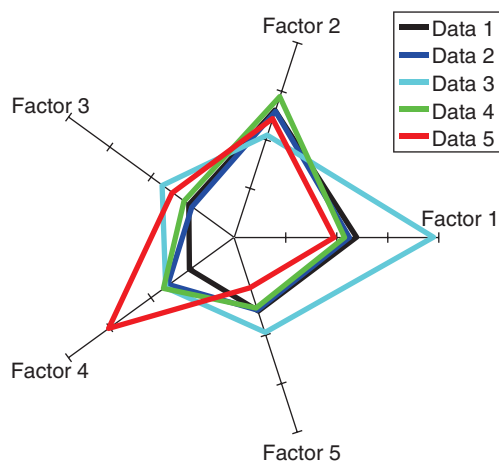


Fig. 15 Direct search results (High-speed search)

Fig. 15, 16の案別解からユーザは、所望の諸元値を選択することが可能である。また、二つのダイレクト探索手法を行うことで、同じ性能値でも、異なる解の候補を得ることが可能である。

3.2. グループ探索

グループ探索は、DBのデータを利用したデータ選好法である。

ユーザは、レーダチャートで性能値を確認しながら諸元値を探索することが可能である。この探索法はクラスタリング技術を用いているために、大規模データの中から、所望の性能値を直接見つけだすことをせずに、クラスデータを利用して、段階的かつ、より簡単にデータを選好することが可能になっている。クラスデータを利用した選好フローをFig. 17に示す。

- (1) DBの10万ケースデータを五つのクラスにクラスタリングし、性能値のレーダチャートを表示する。

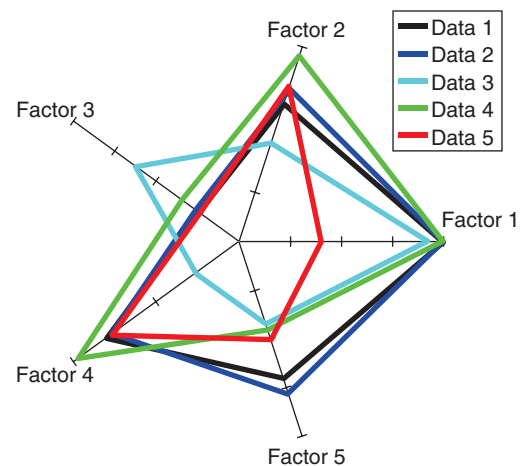


Fig. 16 Direct search results (High-accuracy search)

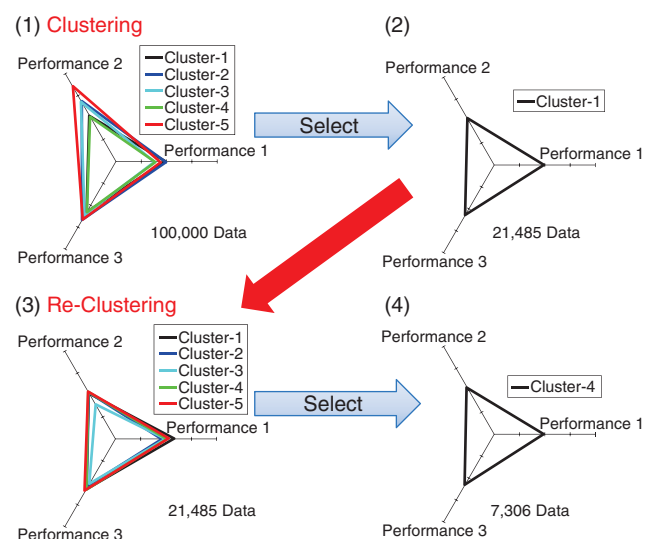


Fig. 17 Decision making flow

- (2) ユーザは所望の性能値で構成されるクラスを選択する。結果として 10 万ケースデータは 21,485 ケースデータに削減される。
- (3) 21,485 ケースデータを五つのクラスに再クラスタリングした性能値のレーダチャートを表現する。
- (4) ユーザは五つの性能値の中から所望のクラスを選択する。結果として 7,306 ケースデータに削減される。さらにケースデータを少なくするように絞り込む作業を続ける。

ここでは代表する五つのクラスデータを用いることで、選好を容易にしておき、選好が進むことで、データの数が減っていることから絞り込みが進んでいることがわかる。

最終的に絞り込んだ結果を Fig. 18 に示す。

Fig. 18 は左図に性能の絞り込み結果を示し、右図に諸元値の案別解を五つまで絞り込んだ結果を示す。

このように、グループ探索は、レーダチャートを活用し、性能間のバランスを確認しながら、データの絞り込みが可能である。

3.3. 寄与度を利用した解析

本システムは、対話型データ選好を使って所望の性能に対する諸元値を知ることができ、ユーザはその諸元値の案別解から納得のいくデータを選ぶことで、従来よりも開発効率を上げることが可能となった。さらに、ユーザは寄与度の分析結果を用いることで、詳細な諸元値セッティングが可能であることを以下に説明する。性能 (Performance1) に対する寄与度の分析結果を Fig. 19 に示す。Fig. 19 は上

図に各諸元の寄与順位、下図に性能に対して五つの諸元で表される部分従属プロットを示している。上図から、性能を向上させるためには変数 2 のパラメータを変更すればよいことがわかる。さらに下図の部分従属プロットの結果を用いることで、パラメータのセッティング方向の理解が可能である。ここでは変数 2 の数値を大きくすることで性能が向上することが示されている。

変数 2 の数値を増加させたときの性能変化を Fig. 20, Fig. 21 に示す。Fig. 20 は五つの諸元から成り立つレーダチャートで、Data2 は Data1 から変数 2 の値を変更した結果である。Fig. 21 は変数 2 の値を変更した時の性能変化を示している。この結果から、Data1 よりも Data2 の方が性能向上したことが示されている。

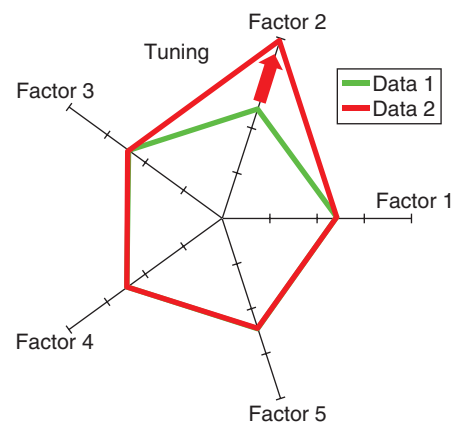


Fig. 20 Factor setting

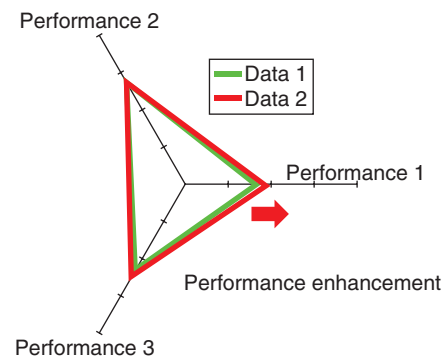


Fig. 21 Performance results

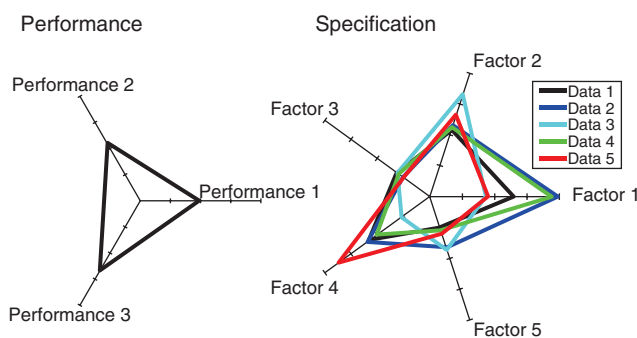


Fig. 18 Decision making results



Fig. 19 Results of Impact analysis

4. まとめ

本研究において、車両開発初期段階で各性能の目標値割り当てが可能な車両性能設計システムを三つのデータマイニング技術で構築した。

- (1) 少数データをもとに高精度な予測を実現する応答曲面生成技術
- (2) ユーザの意思を反映した高効率なデータ探索ができる対話型データ選好技術

- (3) 性能と諸元の関係から、性能に大きく影響する諸元がわかる寄与度分析技術

本システムを車両運動性能シミュレーションに適用したことにより、簡単に分析でき、かつ高速に性能値や諸元値を決定できることを確認した。

参考文献

- (1) 福重貴浩, 有吉斗紀知, 宇田有佑, 大熊友貴, 岡部達哉, 山本幸一, Graening, L., Menzel, S., Olhofer, M.: リアルタイム空力デザインシステムの開発, Honda R&D Technical Review, Vol.24, No.2, p.102-110
- (2) 津野康一, 小谷尚史, 土橋学, 小出文隆, 福重貴浩, 有吉斗紀知: Honda Smart Home System の開発, Honda R&D Technical Review, Vol.26, No.1, p.11-16
- (3) Hastie, T., Tibshirani, R., Freidman, J.: 統計的学習の基礎—データマイニング・推論・予測—, 共立出版, p.422-424, (2014)
- (4) Blatman, G., Sudret, B.: Adaptive sparse polynomial Chaos expansion based on least angle regression, Journal of Computational Physics, Vol. 230, p.2345-2367, (2011)
- (5) Schobi, R., Sudret, B., Wiart, J.: Polynomial Chaos-based Kriging, ETH Zurich Project Report RSUQ-2015-002, February 13th, (2015)
- (6) Bezdek, J. C., Ehrlich, R., Full, W.: FCN: THE FUZZY c-MEANS CLUSTERING ALGORITHM, Computers & Geosciences, vol. 10, p. 2-3, (1984)

■ 著 者 ■



有吉斗紀知



春松正敏



矢野和重



福重貴浩



山田健太郎



鳴川 要