

論文・解説

27 ガウス過程回帰を用いたソフトセンサー技術に関する研究 —エンジン筒内の空気充填効率の推定について—

Estimation of Air Charging Efficiency Using Gaussian Process

小川 史恵^{*1} 桑原 一徳^{*2} 平野 拓男^{*3}
Fumie Ogawa Kazunori Kuwahara Takuo Hirano
目良 貢^{*4} 和田 正義^{*5} 松江 浩太^{*6}
Mitsugu Mera Masayoshi Wada Kota Matsue

要約

自動車エンジンの低排出ガス・低燃費化の鍵を握るのは、エンジン気筒内の空気量を高精度に予測することによる空燃比制御技術の実現である。そのためにはさまざまな運転シーンにおいて、過渡的に複雑に変化する筒内の空気量をリアルタイムに把握することが必要である。筆者らはこの課題を解決するために、統計モデルである Gaussian Process（以下、GP と記す）を用いた推定技術について研究している。本報告では WLTC モード中の条件の下、GP を用いて空気量を予測するための説明変数の選択方法やそれによる推定結果について検討した結果を報告する。

Abstract

To achieve low emission and fuel consumption of vehicle engines, Air Fuel Ratio Control technology is a key technology, which accurately predicts the amount of air in engine cylinders. For the technology, it is necessary to recognize the amount of air in the cylinders in real-time, which transiently changes in a complex way according to driving scenes. We have been studying on estimation technology using a static model called Gaussian Process (GP). Here we report on our study: a selection method of choosing predictor variables for predicting air volume by using GP under WLTC mode, and the estimations.

Key words : Heat engine, Spark ignition engine, Theory/Modeling, Statistic modeling, Gaussian Process

1. はじめに

自動車エンジンの低排出ガス・低燃費化への関心が高まり、燃料をよりクリーンかつ効率的に燃焼する技術が求められている。その鍵を握るのは、エンジン気筒内の空気量を高精度に予測することによる空燃比制御技術の実現である。そのためには、さまざまな運転シーンにおいて過渡的に複雑に変化する筒内の空気量をリアルタイムに把握することが必要である。しかし一般的に、筒内の空気量を推定しようとしてもエンジンの構造上、筒内に空気量センサーを配置することができない。それに対応するために、物理方程式から求めたモデルを通して推定することが考えられるが、物理方程式は過渡的な変化を完全に記述することが困難なことが多い。すなわち実

機の過渡実験データから、エンジン気筒内部において過渡的に変化する空気量の値を高精度に推定する技術の開発が課題である⁽¹⁾⁽²⁾。

筆者らはこの課題を解決するために、統計モデルとして知られる Gaussian Process（以下、GP）を用いた推定技術について研究している。GP は、純粋に統計的な手法と比較して、非線形に強い利点を有している。加えて本技術を発展させることで、最小限のセンサー情報から他の多くの状態量をリアルタイムに推定できるソフトセンサーの実現にも貢献できる可能性がある。

本報告では WLTC モード中の低中速条件を中心に GP を用いて空気量を予測するための説明変数の選択方法や、それによる推定精度（相対誤差）について検討した結果を報告する。

*1,4 技術研究所
Technical Research Center

*2,3,6 パワートレイン開発本部
Powertrain Development Div.

*5 東京理科大学
Tokyo University of Science

2. 実験及び解析方法

2.1 データの取得方法

本節ではモデルとした内燃機関の装置について述べる。装置は排気量 1.997L, 4 気筒の可変バルブタイミング機構を備える直噴ガソリンエンジンを用いた。ガソリンエンジンの空気の流れについて説明する。ここでは、エアクリナーに吸入された空気がスロットルバルブを通じてサージタンク内に流入する。サージタンク内の空気は、クランクの回転に同期して開閉する吸気バルブを通じてシリンダー内に流入する。シリンダー内に流入した空気はシリンダー内に直接噴射される燃料と混合し、火花点火により燃焼する。燃焼後のガスは同じくクランク回転に同期して開閉する排気バルブを通じて排気管に排出され、排気ガスの浄化装置などを通過して大気に出される。

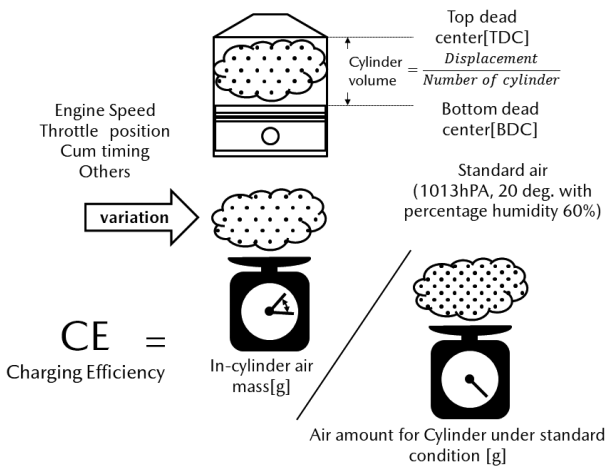


Fig. 1 Charging Efficiency

研究で着目したのは、上記一連の流れの中で筒内へ流入する空気の質量である。一般的に充填効率 (Charging Efficiency: 以下 CE と記す) というパラメーターで表される (Fig. 1 に示す)。CE は筒内に流入した空気の質量を、行程容積 (1.997L/4 気筒=0.499L) を占める標準空気 (20°C, 1atm) の質量で正規化した値である。この CE は一般に直接計測することが難しいため、推定量が用いられ推定には物理モデルや近似物理モデルで行われる。吸気系のダイナミクスを表現した物理統計モデルにおいては、エアフロセンサーやクランク角センサーを用いて計算するエンジン回転数など、多くのセンサー値を使用して各自動車メーカーが工夫を凝らしているところである。従って CE の推定精度を高めることやより少ないセンサー数、更には計算負荷の少ないアルゴリズムで推定が可能になる技術開発が重要である⁽²⁾。

本研究では、エンジンシステム (各センサー含む) 及び 1 次元吸排気シミュレーションツールの解析値を用いている。Table 1 にエンジンの設定条件を記す。

Table 1 Technical Information

Drive system	Front wheel drive 6AT
Engine capacity	1.997L
Engine type	Water-cooled in-line 4cylinder DOHC 16valve engine
Compression ratio	13
Fuel type	Unleaded regular (91RON)

2.2 データの解析手法 (Gaussian Process)

統計モデリングには、ガウス過程回帰を用いた (Gaussian Process, 以下, GP)。GP に関するハイパーパラメーターは、入力する訓練データに基づいて最尤推定を行うことで決定する。ハイパーパラメーターにより、各データに対して時系列的な変化を表現する曲線 (カーネル関数) を使ってモデル化できる。入力データ値によって何通りかの多項式を描けるが、最尤法を用いて 1 つのカーネル関数を使って表現されることが利点である。本手法は Rasmussen (1996) の論じるガウス過程が元になっている⁽³⁾⁽⁴⁾。

以下では GP の概要を述べる。x を説明変数ベクトル, y を目的変数とし, x と y を誤差項 ε とともに結び付けた関数モデル

$$y=f(x) + \varepsilon \tag{1}$$

をベイズ推定することを考える。x と y の n 組のペアからなる訓練用の観測データセット (2) 式がサンプルとして得られたとする。

$$\{(x_i, y_i) \mid i = 1, 2, \dots, n\} \tag{2}$$

行列 (3) とベクトル (4) を定義し、

$$X=[x_1, x_2, \dots, x_n] \tag{3}$$

$$y=[y_1, y_2, \dots, y_n]^T \tag{4}$$

$$\varepsilon_i=y_i-f(x_i) (i=1, 2, \dots, n) \tag{5}$$

誤差 (観測ノイズ) の (5) においては互いに独立で同一の正規分布 (6) 式

$$\varepsilon_i \sim N(0, \sigma_n^2) \tag{6}$$

に従うことを仮定する。また、関数 f(x) の実現値

$$f \mid X=[f(x_1), f(x_2), \dots, f(x_n)] \tag{7}$$

においては適当なカーネル関数 k(x, x') を用いた対称行列 (8)

$$K=k(X, X)=\begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \dots & k(x_n, x_n) \end{bmatrix} \tag{8}$$

に基づき、平均ベクトル $\mathbf{0}$ 、分散共分散行列 \mathbf{K} なる多変量正規分布 (9) 式に従うことを仮定する。

$$f|X \sim N(\mathbf{0}, \mathbf{K}) \tag{9}$$

すると、任意の点 (検証用の説明変数ベクトル) x_* における y の予測分布については式 (10) (11) (12) の形でベイズ推定することができる。

$$y_* | x_*, X, y \sim N(\mu_*, \sigma_*^2), \tag{10}$$

$$\mu_* = k(x_*, X) \cdot (K + \sigma_n^2 I)^{-1} y, \tag{11}$$

$$\sigma_*^2 = k(x_*, x_*) - k(x_*, X) \cdot (K + \sigma_n^2 I)^{-1} k(X, x_*) \tag{12}$$

ここに、 μ_* は予測平均、 σ_*^2 は予測分散、 $k(x_*, X)$ は

$$k(x_*, X) = [k(x_*, x_1), \dots, k(x_*, x_n)] = k(X, x_*)^T \tag{13}$$

を表す。上記の μ_* が特に y の x_* における (平均的な) 推定値を与えており、

$$\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T = (K + \sigma_n^2 I)^{-1} \tag{14}$$

とおけば (15) 式のように表すことができる。すなわちカーネル関数の線形結合であることを示している。

$$\mu_* = \sum_{i=1}^n \alpha_i \cdot k(x_*, x_i) \tag{15}$$

以上を踏まえ、ここでは関数 $f(x)$ を具体的に

$$f(x) = \sum_{r=1}^m \{a_r \cos(2\pi s_r^T x) + b_r \sin(2\pi s_r^T x)\} \tag{16}$$

(ただし、係数ベクトル s_r や基底関数の個数 m はハイパーパラメーター) なる特別な形として取扱うことにし、計 $2m$ 個ある a_r, b_r ($r=1, 2, \dots, m$) については互いに独立で同一の正規分布 (17)

$$a_r, b_r \sim N\left(0, \frac{\sigma_*^2}{m}\right) \tag{17}$$

に従った乱数を発生させることにより定義する。実際このように取扱えば $f|X$ は平均ベクトルが $\mathbf{0}$ であり、分散共分散行列がカーネル関数 (18)

$$k(x_i, x_j) = \frac{\sigma_*^2}{m} \sum_{r=1}^m \cos(2\pi s_r^T (x_i - x_j)) \tag{18}$$

を共分散にもつ \mathbf{K} であるような多変量正規分布に従う一例として構成することができ、上記に述べたとおり、任意の点 x_* における y の推定値を μ_* として得ることができる。Table 2 に計算条件を示した。

Table 2 GP Simulation Parameters

Name	Parameter Setting
Kernel as subset size (m)	50
Number of Iteration	100
Minimum Time Lag Inputs/Outputs	0, 1
Max Time Lag Inputs/Outputs	2/2, 5/5, 10/10
Kernel function	Trigonometric

2.3 検証内容

一般的に GP を用いた過渡モデリングは①データ収集、②データ前処理、③モデル構築、④モデル検証のプロセスで開発される。本報告では特にモデルの精度に影響が大きいと考えられる①②のプロセスに着目し目的関数を表現する時系列データの加工方法と、説明変数の選択方法について検証する。

3. 結果

3.1 時系列データ解析における諸検証

(1) データの分割

GP を用いた検証に当たり、Fig. 2 のように合計 600 秒間からなる車両の走行データに対し、半分の 300 秒を訓練データ、残りを検証データと割り当てた。訓練データに対して分割周期の変更を行い、検証データを入力し検証を行った。

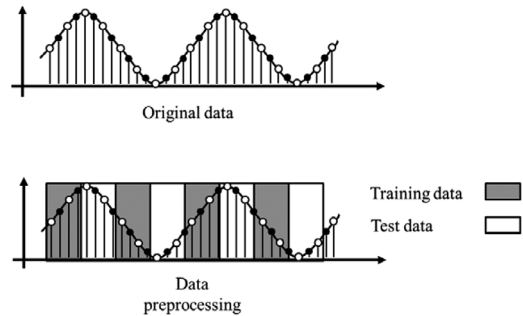


Fig. 2 Conceptual Diagram of Preprocessing

(2) Lag についての検討

本検証で使用しているような時系列データの性質をつかむためには、統計量の算出が重要である。統計量の指標としては、自己相関関数や偏自己相関関数がある。これらは時系列データにおける各時刻のデータが、時間差を設けてずらしたデータとどれだけ相関関係にあるかを定量化する指標である。時系列データでは観測値と観測した時間が逐次記録されており、各観測した時点の関係性 (前後関係) に意味がある。本研究では時系列データのもつ時間差の度合いである Lag (ラグ) に注視し、時点をずらして算出した自己相関関数を用い検証の指標とした。

なお GP 回帰に使用したソフトウェア ASCMO では Lag について仮説を立てやすくするため、Fig. 3 に示すように逆自己相関関数 IACF (Inverse Autocorrelation Function) と呼ばれる指標を参照している⁽⁵⁾。横軸に時間の Lag、縦軸に IACF を示しておりゼロに収束していく IACF の出力結果を参照しつつデータごとの Lag を検討するのが望ましいとされている。本結果より、Lag が 0, 1 を中心に相関を持ち 2 以上の場合でも影響を受けていることが示唆される。そのため Lag をここでは大まかに 0, 2, 10 と仮定して後節で検証した⁽⁶⁾。

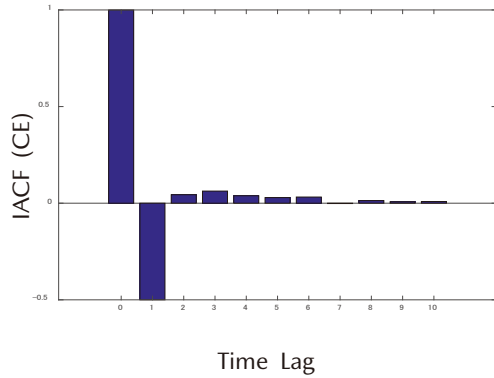


Fig. 3 Result of Inverse Autocorrelation Function

3.2 説明変数の選択方法に関する検討

説明変数の選び方について検討した結果を示す。推定する目的変数はCE, 真値はECUで得られるCE値とした。説明変数は(a)~(c)の考え方に従い選択し, 後述する k-fold cross validation 法を用いてモデル精度を比較検証した。

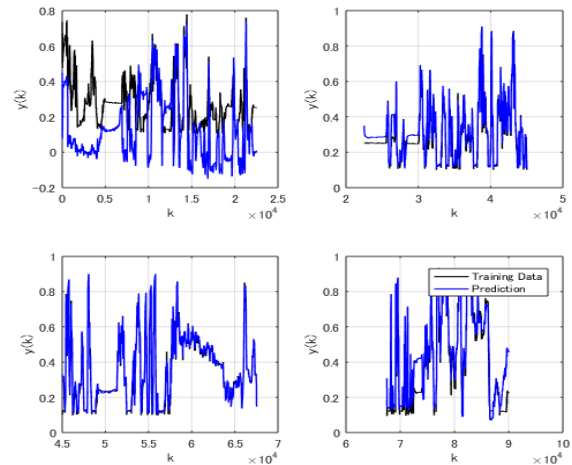
- (a) 物理的センサーから取得できる全ての因子 (気筒近傍の全てのセンサー値) を説明変数にした場合
- (b) 機械の状態に関するセンシング値と, CEとの因果が強い温度に関連する因子のみを説明変数とした場合: エンジン回転数, スロットル開度, カムタイミング (インテークバルブ/エキゾーストバルブ)+温度
- (c) 機械の状態に関するセンシング値のみを説明変数とした場合: エンジン回転数, スロットル開度, カムタイミング (インテークバルブ/エキゾーストバルブ)

適合度指標を Table 3 に, 解析結果を Fig. 4 にそれぞれ示す。Fig. 4 の黒線は訓練データ, 青線は推定値である。検証には k-fold cross validation (k 分割交差検証) を用いた。これは k セットに分割されたデータ内の 1 セットを除いたデータで学習して, 回帰モデルを構築して残りの 1 セットで検証するものである。本検討では分割数 k = 4 とした。すなわち, 元の時系列を計 4 つの区間に分割して取扱うことを意味する。結果は全て RMSE が 0 に近い場合をよい変数の選び方として評価する。

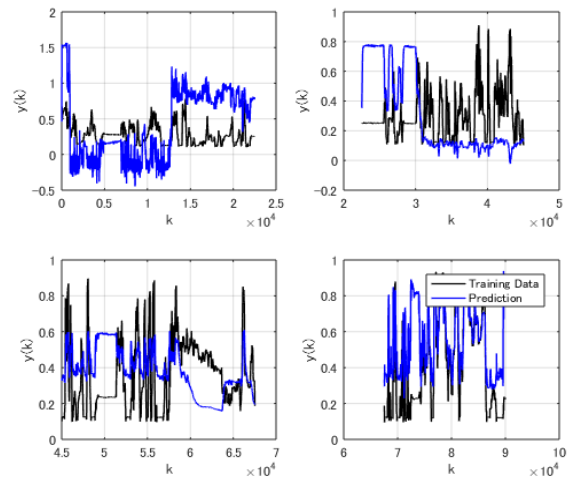
Table 3 Output of 4-fold Cross Validation

(a)	
RMSE: 0.23085 R ² : -1.70510	RMSE: 0.03145 R ² : 0.98071
RMSE: 0.02273 R ² : 0.98536	RMSE: 0.07971 R ² : 0.91062
(b)	
RMSE: 0.48899 R ² : -11.1373	RMSE: 0.34740 R ² : -3.79630
RMSE: 0.21922 R ² : -0.36169	RMSE: 0.23634 R ² : 0.21436

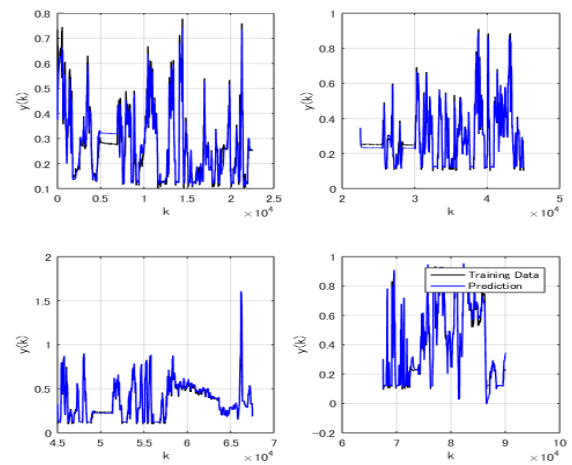
RMSE: 0.02326 R ² : 0.97254	RMSE: 0.01459 R ² : 0.99154
RMSE: 0.05716 R ² : 0.90742	RMSE: 0.03863 R ² : 0.97901



(a) For All Parameters Modeling



(b) For Thermal Parameters Modeling



(c) For Vehicle Speed, Throttle Position and Cam Timing Modeling

Fig. 4 Output of 4-fold Cross Validation

上記の結果から特に (c) のケースにおいて、全ての検証データで R^2 が高く確保され統計モデルの説明変数の選択として妥当と判断した。また、GP では多くの説明変数を与えると逆に推定精度が低下することを把握した。

(a) や (b) での推定精度が低い理由として、(a) のケースでは過度に多くの変数を含むため、逆行列計算に特異な影響が出てしまい推定精度が低下したことが考えられる。(b) については (c) の変数に追加して温度の変数を更に加えることで計算したが (c) の変数の中で温度との相関関係が生じたことによる影響で、(b) のケースでは多重共線性が排除できなくなったと考える。

3.3 遅れ要素を考慮したフィルター処理の検討

一般的に内燃機関における制御は、エアフロセンサー (AFS) をエンジンの吸気系上流に配置することで気筒に充填される大まかな空気量を計測できる。ただ、気筒内内部や近傍にセンサーを配置できないことを先に述べた。そのため、これまでの経験から CE を正確に把握するには AFS から気筒までの吸気経路分の遅れ要素を考慮するとよいとされてきた。

そのため本節では、遅れ要素を表現できる LPF (ローパスフィルター) やダウンサンプリングを検討した。なかでもダウンサンプリングは LPF のような特性をもつことが知られている。

Fig. 5 では Lag=2 の場合の CE の推定値を示した。

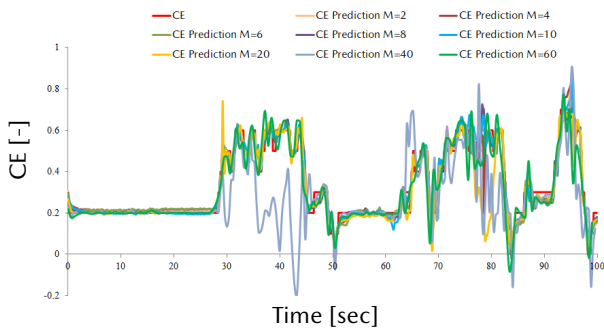


Fig. 5 Comparison of Prediction for Down Sampling with Lag=2

ダウンサンプリングの方法としてはデシメーション (間引き) を実行する。目的変数 (CE) を表す y に関するダウンサンプリング前の信号を $y[t]$ 、後の信号を $y[M \cdot t]$ とする。 M はダウンサンプリング比と呼ぶ。 $y[t]$ の実際のデータ $y_i (i = 1, 2, \dots, n)$ は各時刻 $t = i \Delta t$ (Δt は時間刻み幅 [sec] を表し、本検討では $\Delta t = 0.01$ である) における値をもち、ダウンサンプリングにより得られる $y[M \cdot t]$ の実際のデータ $y_{i'} (i' = 1, 2, \dots, n')$ は各時刻 $t = i' M \Delta t$ ($M \Delta t$ が時間刻み幅 [sec]) における値をもつことになる。信号を間引くことで時系列データのもつ周波数スペクトル成分が変化する。つまり一部の信号列がゼロ値に置き換わり、従来は生じていないデータ上にもノ

イズが生じることがある。 M が低下すると、隣に接するスペクトル同士が近づく。 M が大きくなるとエイリアシングが起きる傾向もみられる⁽⁷⁾。そのためダウンサンプリング比を選択して用いることは、真値に対する推定において効果的と考えられる。Fig. 6 では、 $M=40$ (すなわち、 $M \Delta t = 0.4$ [sec] 刻み) に設定すると推定精度が低下した。この要因は隣接する元のデータ同士が相互に影響したためと考えられる。説明変数と目的変数それぞれでスペクトルの変化が伴い推定精度に影響が出るため、適切な M の値を選ぶ必要がある。今回は Lag=2 のみの結果であるので、次節では Lag の条件を変えて検討を行うこととした。

3.4 Lag の変化に伴う統計モデルの推定精度

本節では CE の推定を Table 1 のエンジン諸元及び Table 2 の計算条件の下で、Lag 違いの解析を行った。結果を Fig. 6 へ示す。

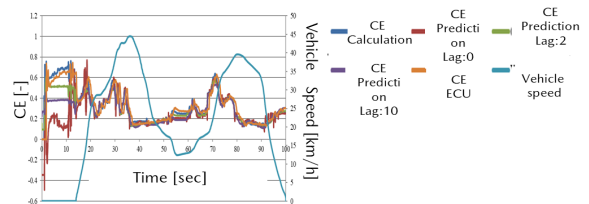


Fig. 6 Comparison of Prediction for Some Lag Values

同図では Lag=0, 2, 10 を比較して縦軸の左側が CE、右側が車速を示している。CE 計算値 (青) は AFS 取得の流量瞬間値と CAE で計算した CE から推測した値でありこれを真値とした。ECU による出力値である CE (橙) と比較すると、GP で推定した CE に比べて若干の時間遅れが生じている。GP による推定値と各 Lag を比較すると横軸の時刻 20 [sec] 以上から 50 [sec] までは、どの Lag においても CE 値が推定値とほぼ一致していた。ただし比較的急峻な加速シーンにおいて、Lag の値を 10 で推定させると再現性が低いことがわかった。一方で Lag を 0 とした場合では、走行開始してから若干の上限振動はあるがほぼ追従して真値に近い推定値を保っていた。

3.5 統計モデルの活用法に関する考察

以上の結果から下記のことが考察できる。

走行パターンを考慮して内燃機関を併用させる車両の場合 Lag が大きい場合には CE 値の推定精度が低くなるため、急峻な加速シーンにおいては IG-ON (エンジン始動) と連携させて考える必要がある。例えば制御の ON が入った後、50ms (0.05sec) ~ 1sec 以降にランプ (始動までの勾配部分のこと) なしで推定開始するのではなく、計算のための予備の時間としてランプ期間を設けるのが効率的と考える。

また、燃料の濃度の配慮も必要であると考えられる。例え

ば理論空燃比からの逸脱にある程度の閾値などを設け使用することが妥当だと考えるが、フューエルカット復帰時や吸入空気量が少ない時は更に注意が必要である。その理由は、フューエルカット中は筒内の既燃ガスが空気に置換するために空気量が通常運転と数十%程度異なり、また吸入空気量が少ない場面では空気量の予実差によりそれぞれA/Fが理論空燃比から大きく逸脱してエンジンは失火してしまうことになるからである。従って、燃料の濃度についてはリーンかつリッチであっても閾値を設けてモデルを動作させるのがよい。

更には、Lagを0以外に設定することだけでなく、車速が一定とならない領域を十分長く確保できるシーンにおいてLagの変化も考慮しつつGPによる統計モデルを学習させることが重要であると考えられる。これらを踏まえ、Fig. 3に示す指数関数的に最小値を示したLag=10とそのほかのLag値の設定を最適化しながらGPモデルを活用すれば、安定した制御を求めることができると考える。

以上のように統計モデルの推定精度は、説明変数の選択やデータの前処理などに大きく影響を受ける。これらを適切に設定しながらモデルの精度を正しく把握し、制御モデルへの反映可否を検討していくことが重要である。特に燃料のリーン状態などでは、推定精度を出すのが難しく、GPモデルの適用が限定的となる可能性も考えられる。

4. まとめ

本研究ではエンジン気筒内の空気量を高精度に予測するため、統計モデルであるGPの適用方法を検討し、説明変数の選び方や時系列データのLag設定がモデル精度に対して重要となることを述べた。本検討結果はエンジンを対象とした統計モデリングの手法の一つとして有効であり、モデル作成のためのデータ群を学習させる上では車両の運動状況も重要であることがわかった。

今後も更なる低排出ガス・低燃費化の実現に向け、統計モデルによるCE推定精度を更に向上させる。それと並行して、広く他の課題にも適用できるように、本技術を基に汎用的な統計モデル化の技術基盤に発展させていきたい。

参考文献

- (1) R. Benedikt: Local Gaussian Process Regression in order to Model Air Charge of Turbocharged, Gasoline S5-Engines, 2016-01-0624, 04/05/2016, BMW (2016)
- (2) 船津, 金子: ソフトセンサー入門—基礎から実用的研究例まで—, (2014)
- (3) Christopher Bishop: Pattern Recognition and Machine Learning, Springer (2006)
- (4) Carl Edward Rasmussen and Christopher K. I. Williams: Gaussian Processes for Machine Learning, MIT-Press (2006)
- (5) Ette Harrison Etuk: On Autogressive Model Identification, Journal of official Statistics, Vol.4, No.2, pp.113-124 (1988)
- (6) 小川, 清水, 和田: ガウス過程を用いたガソリンエンジンの吸入空気量推定における統計モデリング手法の開発, 自動車技術会論文集, 2019年50巻2号, pp.260-266 (2019)
- (7) 足立: ユーザのためのシステム同定理論, コロナ社 (1993)

付録

R^2 (Coefficient of Determination)

モデルの誤差を評価するための相対的な指標、全部の測定データの平均値に対する分散とモデルに対する分散の比を示す。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{i,predictive} - y_{i,true})^2}{\sum_{i=1}^n (y_{i,measure} - \bar{y}_{i,true})^2}$$

$y_{i,predictive}$: 推定値, $y_{i,true}$: 真値

$\bar{y}_{i,true}$: 真値の平均, n : 測定データの数

RMSE (Root Mean Square Error)

平均化された誤差の値を表す。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{i,predictive} - y_{i,true})^2}{n}}$$

■ 著 者 ■



小川 史恵



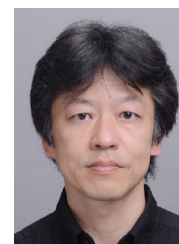
桑原 一徳



平野 拓男



目良 貢



和田 正義



松江 浩太