2002年日本オペレーションズ・リサーチ学会
春季研究発表会

サポートベクターの感度を考慮した追加学習

甲南大学　＊鷲野　宏治　(WASHINO Koji)
甲南大学　中山　弘隆　(NAKAYAMA Hirotaka)

1　はじめに

実際の工学設計の問題では目的関数が設計変数の陽な関数として表すことができず、設計変数の値を決めたとき、適当な解析 (構造解析、流体解析、熱解析など) あるいはシミュレーションによって初めて目的関数の値が求まることが多い。通常、解析等のシミュレーションによって未知目的関数 $f(x)$ の値を求めるには多大の時間を要する。したがって、多くの目的関数評価回数を要する従来の最適化手法をそのまま用いれば、解を得るまでに膨大な時間を要することになる。そこで、本論文ではいくつかのサンプル点から関数 f の形を学習させ、f に対して最適化を行い最適解を予測するという繰り返しによって解を求めるという方法について論ずる。

以下、目的関数の近似には Support Vector Machine(SVM) を用い、近似目的関数に対する最適化には遺伝アルゴリズムを用いた手法を紹介する。

2　Support Vector Machine(SVM) を使った関数近似

SVM は主にパターン分類に用いられるが、Linear ε-insensitive loss function

$$L^{\varepsilon}(x,y,f) = |y - f(x)|_{\varepsilon} = max\ (0, |y - f(x)| - \varepsilon)$$

を導入することで関数近似にも応用できる。Linear ε-insensitive loss function とは、学習データ y_i と近似関数 $f(x_i) = w^t x_i + b$ の差の絶対値をある値 ε より小さくなるように $f(x)$ を近似する。原空間に与えられた学習データ (x_i, y_i) $(i = 1, ..., l)$ に何らかの非線形写像 $\phi : R^n \to R^q$ を行うことにより高次元特徴空間に写像し、近似関数 $f(x) = w^t \phi(x) + b$ を考える。高次元特徴空間に写像することにより、$|y - f(x)| < \varepsilon$ となることが期待されるが必ずしも保証されるものでないためスラック変数 ξ_i $(i = 1, ..., l)$ を導入する。SVM における関数近似は $|y_i - f(x_i)| < \varepsilon + \xi_i$ $(i = 1, ..., l)$ を制約にもつ 2 次計画法を解くことに帰着される。

2 次計画法主問題の制約に対するラグランジェ乗数を $\alpha, \acute{\alpha}$ とし双対問題を考える、目的関数に高次元ベクトルの内積 $\phi(x_i)^t \phi(x_j)$ が出てくるが、原空間の距離を像空間の内積として保存するカーネル関数 $K(x_i, x_j) = \phi(x_i)^t \phi(x_j)$ を使うことにより、高次元ベクトルの内積の計算をすることなくカーネルの値を求めるだけで、簡単に最適化を行うことができる。このようなカーネルトリックを使うことにより、近似関数は $f(x) = \sum_{i=1}^{l} (\acute{\alpha}_i - \alpha_i) * K(x_i, x) + b$ で与えられ、後述の 2 次計画法双対問題の解 $\alpha_i^* = \acute{\alpha}_i - \alpha_i$ により、関数の形を近似できる。

$\alpha_i^* \neq 0$ の時 i 番目学習データを Support Vector と呼び、関数近似する際に $|\alpha_i^*|$ が大きいほど i 番目の学習データに対する $K(x_i, x)$ の値を近似関数 $f(x)$ に大きな影響を与えるため重要な学習データとして考えられる。学習データを追加する際 Support Vector の情報、つまり 2 次計画法主問題の制約の感度 α_i^* を考慮する。

3　近似目的関数に対する最適化のための遺伝アルゴリズム

上述の方法によって近似された目的関数に対する最適化を行うには、すでに近似目的関数は設計変数によって陽に表されているので、遺伝アルゴリズムによって最適値を求める。遺伝アルゴリズムは設計変数が離散値の問題に有効であるが、本研究ではシミュレーションとして連続値を使用しているために、連続変数に対する遺伝アルゴリズムとして BLX-α を使用している。この方法は遺伝子のコード化に対し、通常の遺伝アルゴリズムでよく用いられるバイナリーコーディングあるいは、グレイコーディングを採用せず、実数値のまま用い、交叉において両親を対角にもつブロック内で、ランダムに子を生成するのが特徴である。

4 追加学習点の選択

本研究では、目的関数の評価回数をできる限り少なくするために少なめの学習データから出発し、十分な近似最適解が得られるまで少しづつ学習データを追加し、効果的な追加学習を行うことより目的関数の予測精度を向上させることが課題である。

そこで、より精度の高い近似最適解を求めるために局所的情報として、GA によって求められた最適点の近傍に1つ目の追加学習データを追加する。また、極値への収束を防ぐために大局的情報として2つ目の追加学習データを追加する。

学習データの追加手法として、次のような方法を提案する。ランダムに学習データの候補 20 点を発生させ、既存の学習データとの密度を考慮し密度の低い学習データの候補 $\boldsymbol{x_c}$ に対する変数、制約を 2 次計画法双対問題に追加し、解きなおす。簡略化のために変数と制約を追加した 2 次計画法主問題を下に示す。この時、学習データの候補 $\boldsymbol{x_c}$ に対する学習データを y_c ではなく、関数の近似値 $f_{predict}(\boldsymbol{x_c})$ を使う。追加した学習データの候補に対する $|\alpha_c^*|$ と他の $|\alpha_i^*|(i = 1, ..., l)$ に対する相対的な大きさを考慮する。この操作を密度の低い学習データの候補 5 点に対し行い、相対的に $|\alpha_c^*|$ が 1 番大きな学習データの候補を追加学習データとして追加する。

学習データの追加の方法として局所的情報、大局的情報ともに上述の方法で行うが、ランダムに学習データの候補を発生される範囲が、局所的情報の場合は最適点に収束すればするほど最適点の近傍で小さくなるような範囲で発生させるのに対し、大局的情報の場合は学習データの範囲全体で学習データの候補を発生させる。

2 次計画法の双対問題	学習データの候補の追加 $\boldsymbol{x_c}$
目的： $-\frac{1}{2}\sum_{i,j=1}^{l} K(\boldsymbol{x_i},\boldsymbol{x_j})(\acute{\alpha}_i-\alpha_i)(\acute{\alpha}_j-\alpha_j)$ $-\varepsilon\sum_{i=1}^{l}(\acute{\alpha}_i+\alpha_i)+\sum_{i=1}^{l} y_i(\acute{\alpha}_i-\alpha_i) \rightarrow \ MAX$	目的： $\frac{1}{2}(\boldsymbol{w}\cdot\boldsymbol{w})+C\sum_{i=1}^{l+1}(\xi_i+\acute{\xi}_i) \ \rightarrow \ MIN$
制約： $\begin{cases} \sum_{i=1}^{l}(\acute{\alpha}_i-\alpha_i)=0 \\ 0 \leq \acute{\alpha}_i \ , \ \alpha_i \leq C \\ (i=1,...,l) \end{cases}$	制約： $\begin{cases} y_i-\sum_{i=1}^{l+1} w_i * \phi(\boldsymbol{x_i}) \leq \varepsilon+\acute{\xi}_i \\ \sum_{i=1}^{l+1} w_i * \phi(\boldsymbol{x_i})-y_i \leq \varepsilon+\xi_i \\ f_{predict}(\boldsymbol{x_c})-\sum_{i=1}^{l+1} w_i * \phi(\boldsymbol{x_i}) \leq \varepsilon+\acute{\xi}_i \\ \sum_{i=1}^{l+1} w_i * \phi(\boldsymbol{x_i})-f_{predict}(\boldsymbol{x_c}) \leq \varepsilon+\xi_i \\ \xi_i \ , \ \acute{\xi}_i > 0 \ (i=1,...,l+1) \end{cases}$

5 終わりに

目的関数の評価回数をできる限り少なくするために、効果的な追加学習データを選び方が本研究の目的でありSupport Vector の情報を使うことにより、これを実現しようとした。本論文では Simulation の結果を示すことができなっかたが $y=-\sin x+\sin\frac{2}{3}x \quad (0 \leq x \leq 20)$ を未知目的関数とし、初期学習データ数:3 とし最大値を求めた結果、学習データ数 15 点で十分精度の高い実行結果が得ることができ、ある一定の成果を得られたのではないかと考えられる。しかし、Support Vector の情報、つまり主問題の制約の感度の情報を使うことによりさらに効果的な学習データを追加できると考えている。また、2 次計画法から線形計画法に変えることにより、計算または感度解析の簡略化、計算時間短縮などが可能である。

参考文献

[1] Nello Cristianini, John Shawe-Taylor：An Introduction to Support Vector Machines and other Kernel-based learning method, Cambridge, 2000

[2] Hirotaka Nakayama, Masao Arakawa, Rie Sasaki : Optimization with Implicitly Known Objective Function Using RBF Networks and Genetic Algorithms, Proc. of ICANNGA, pp.387-390, 2001