Artificial Bee Colony アルゴリズムによるサポートベクターマシンのハイパーパラメータ最適化

2131007 安達 拓真

千葉工業大学 情報科学部 情報工学科 4 年 山口研究室

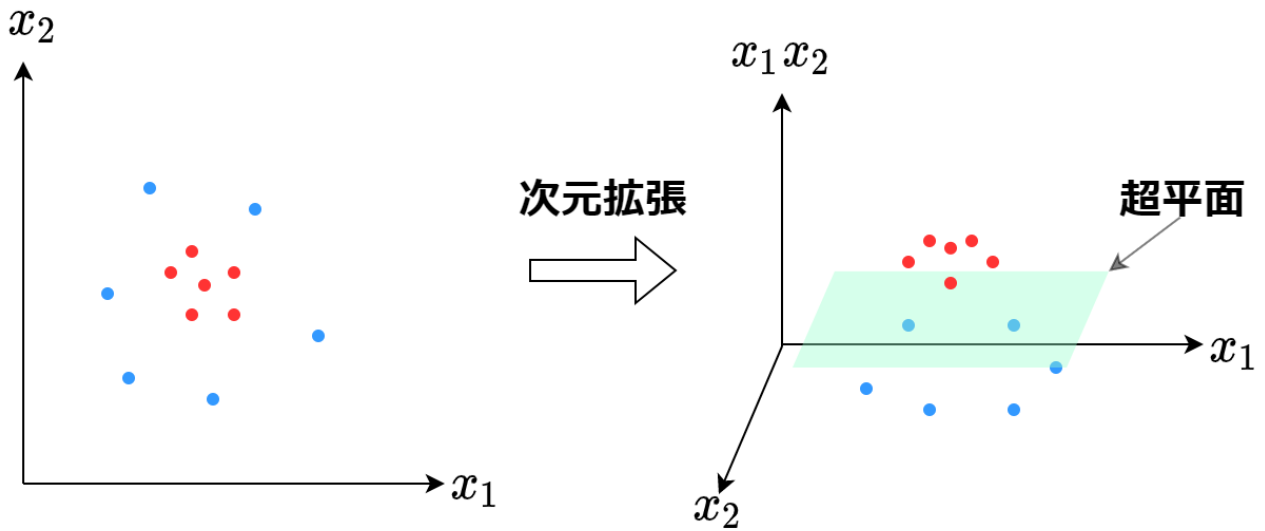
2024 年 1 月 24 日

はじめに

- 機械学習にとってハイパーパラメータはモデルの性能を決める重要な値
- ハイパーパラメータを自動で調整する研究が行われている
- 先行研究として，Artificial Bee Colony(ABC) アルゴリズムを用いて，サポートベクターマシン (SVM) のハイパーパラメータ最適化と特徴選択を行った研究がある
- 本研究では，先行研究で最適化対象ではなかったカーネル関数をハイパーパラメータとして扱う手法を提案する

サポートベクターマシン (SVM)

- 1995 年に提案された，分類や回帰に使用される機械学習アルゴリズム [1]
- 非線形データを高次元空間に写像し，線形分離可能にする
- データを分類する最適な境界線 (超平面) を探す



[1] Cortes, C. and Vapnik, V. Support-vector networks, Ma-chine Learning, Vol.20, No.3, pp.273-297, 1995.

Artificial Bee Colony(ABC) アルゴリズム

- 蜂の採餌行動に着目した最適化アルゴリズム [2]
- 働き蜂，追従蜂，偵察蜂の三種類の蜂によって各食物源の探索を行い，最適解を求める
- 最適化対象は実数値
- ABC 自体の設定パラメータは少ない

[2] Karaboga, Dervis. An idea based on honey bee swarm for numerical optimization. Vol. 200. Technical report-tr06, Erciyes university, engineering faculty, computer engineering department, 2005.

先行研究[3]におけるSVMのハイパーパラメータ最適化

- カーネル関数をRBFカーネルに固定
- 最適化にはABCを使用
- 最適化したハイパーパラメータ
  - SVMのC
  - RBFカーネルのγ

[3]近藤 久，浅沼 由馬“人工蜂コロニーアルゴリズムによるランダムフォレストとサポートベクトルマシンのハイパーパラメータ最適化と特徴選択”，人工知能学会論文誌, vol34-2, pp.1-11, 2019.

問題点

- カーネル関数を RBF カーネルに固定している
  - SVM には RBF カーネル以外にも様々なカーネル関数が適用できる
  - カーネル関数によってハイパーパラメータが異なる
- ハイパーパラメータ空間の探索範囲が限定的

提案手法

- 以下の 4 つのカーネル関数とそのハイパーパラメータも最適化対象とする

$$
\begin{aligned}
\text{線形カーネル}: &\quad K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{x_i}^T \cdot \boldsymbol{x_j} \\
\text{RBF カーネル}: &\quad K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp\left(-\gamma_0 \|\boldsymbol{x_i} - \boldsymbol{x_j}\|^2\right) \\
\text{シグモイドカーネル}: &\quad K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \tanh(\gamma_1 \boldsymbol{x_i}^T \cdot \boldsymbol{x_j} + \mathsf{coef0}_0) \\
\text{多項式カーネル}: &\quad K(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\gamma_2 \boldsymbol{x_i}^T \cdot \boldsymbol{x_j} + \mathsf{coef0}_1)^d
\end{aligned}
$$

カーネル関数が持つハイパーパラメータの扱い

- ABC でハイパーパラメータが異なるカーネル関数を同時に扱う必要がある
- 同じ性質のハイパーパラメータが存在することに着目
  - 4 つのカーネル関数のハイパーパラメータの合計は 6 個
  - 性質ごとに分けると 3 個
- 他のカーネル関数で使用するパラメータの値をそのまま使用
  - ランダム性により解の多様性が生まれる

食物源の形とハイパーパラメータの扱い

- ABC における食物源は 5 個の数値で表す

食物源の形

変数	カーネル関数	C	γ	coef0	d
型	文字列	実数	実数	実数	整数
数値	{1,2,3,4}	[0,1]	[0,1]	[0,1]	[0,1]

- C,γ,coef0,d は以下の式によって SVM に適用される値に変換される
  - 整数である d は四捨五入を行う

$$A = a(a_{max} - a_{min}) + a_{min}$$

- カーネル関数によって異なるハイパーパラメータはカーネル関数の値によって活性，非活性となる

カーネル関数の扱い

- カーネル関数は文字列であるため ABC の更新式が適用できない
- カーネル関数の更新はランダムに選ばれた個体とのルーレット選択

$$P = \frac{\mathrm{fit}(\boldsymbol{x_j})}{\mathrm{fit}(\boldsymbol{x_i}) + \mathrm{fit}(\boldsymbol{x_j})}$$

i: 更新個体　j: ランダムに選ばれた値

実験

- 侵入検知問題である KDD'99 データセットを，
  デフォルトパラメータ，既存手法，提案手法で解く
- 既存手法，提案手法は 10 回ずつ実行し，平均値をとる
- データセットはランダムに 10%抽出した物を 3 つ使用する
  - 学習セット： SVM の学習に使用
  - 検証セット： SVM の評価に使用
  - テストセット： 最終的に得られた最良解の評価に使用

実験パラメータ

ABC の実験パラメータ

パラメータ	値
コロニーサイズ	20
LIMIT	100
サイクル数	500

SVM の実験パラメータ

パラメータ	値
kernel	[linear, RBF, sigmoid, poly]
C	[10^{-6}, 35000]
γ	[10^{-6}, 32]
coef0	[0, 10]
d	[1, 3]

実験結果 (分類精度と実行時間)

- 提案手法はデフォルトパラメータ，既存手法よりも分類精度が高くなった．
- 実行時間は既存手法よりも長くなった．

分類精度と実行時間

	線形	RBF	シグモイド	多項式	既存手法	提案手法
分類精度 [%]	99.68	99.78	96.12	99.76	99.88	99.91
実行時間 [h]	-	-	-	-	11.8	15.5

評価指標

- モデルの評価指標として検知率，誤警報率，適合率，F 値を使用する

混同行列

予測クラス＼実際のクラス	攻撃	通常
攻撃	TP	FP
通常	FN	TN

$$\text{検知率} = \frac{TP}{TP + FN}, \quad \text{誤警報率} = \frac{FP}{TN + FP}$$

$$\text{適合率} = \frac{TP}{TP + FP}, \quad \text{F 値} = \frac{2*\text{検知率}*\text{適合率}}{\text{検知率} + \text{適合率}}$$

実験結果 (評価指標)

- 提案手法では TP，TN が向上し，FP，FN は減少したため，検知率，適合率，F 値が向上し，誤警報率は減少した
  - 侵入検知問題におけるモデルの性能が向上した

混同行列の値

	先行研究	提案手法
TP	39602.4	39609.7
TN	9743.8	9748.9
FP	22.2	17.1
FN	33.6	26.3

モデルの評価指標

	先行研究	提案手法
検知率 [%]	99.91	99.93
誤警報率 [%]	0.23	0.17
適合率 [%]	99.94	99.96
F 値 [%]	99.93	99.95

データの有意性

- 既存手法と提案手法の実験結果の有意性を，t 検定により検証した
  - p 値：既存手法と提案手法のデータに差がないと仮定した時に，本実験の結果が起こる確率
- すべての評価指標で p 値が 0.05 未満のため，有意水準 5%で有意差ありと言える

評価指標ごとの p 値

	p 値
分類精度	0.00020
実行時間	0.00061
検知率	0.028
誤警報率	0.0076
適合率	0.013
F 値	0.0025

考察

- 提案手法で実行時間が長くなってしまった原因
  - カーネル関数をハイパーパラメータとして扱い探索範囲を広げたこと
- 既存手法では RBF カーネルのみを使用
- RBF カーネルは汎用性が高く，他のカーネル関数に比べて学習時間が短い傾向にある
  - RBF カーネル以外のカーネル関数の個体の評価に時間がかかった可能性

おわりに

- カーネル関数もハイパーパラメータとして扱い，SVM のハイパーパラメータを最適化する手法を提案
- 提案手法は先行研究よりも分類精度が高くなったが，実行時間は長くなった
- 探索範囲を広げたことで，データセットに応じた柔軟なモデル構築が可能となる
  - 様々なデータセットで提案手法の汎用性を検証する必要がある