

2024年度

卒業論文



論文題目

Artificial Bee Colony アルゴリズムによる
サポートベクトルマシンの
ハイパーパラメータ最適化

研究者

2131007 安達 拓真

指導教員

山口 智 教授

2024年1月xx日

目次

1	はじめに	1
2	サポートベクトルマシン (SVM)	2
2.1	ハードマージン SVM	2
2.2	ソフトマージン SVM	5
2.3	カーネルトリック	7
2.4	多クラス分類への拡張	9
3	ABC アルゴリズム	10
3.1	概要	10
3.2	探索手順	10
4	提案手法	13
4.1	ABC アルゴリズムの適用	13
4.2	提案手法のアルゴリズム	14
5	実験	17
5.1	データセット	17
5.2	実験設定	18
5.2.1	実験環境	18
5.2.2	パラメータ設定	18
5.3	実験結果	19
6	考察	20
7	おわりに	21

謝辞	22
参考文献	23

1 はじめに

はじめに abc.svm,hpo, 郡知能. ・背景 SVM とはこういうものである。

機械学習モデルにはあらかじめ決めておかなければいけない値であるハイパーパラメータが存在する。そして、機械学習モデルの性能を最大限発揮するためには、適切なハイパーパラメータの選択が必要不可欠である [1]。ハイパーパラメータにはこれがある。SVM のパラメータはそれぞれこういう特徴がある。

一般にハイパーパラメータ調整は手動やグリッドサーチ、ランダムサーチで行われる。手動でハイパーパラメータを調節することは直感と経験に頼る作業になる。グリッドサーチ、ランダムサーチは探索効率が悪い。ハイパーパラメータ調整 (Hyper Parameter Optimization(HPO)) の分野では、離散、連続、カテゴリなど様々なハイパーパラメータを扱い、広い探索空間を効率的に探索することが求められる。HPO では目的関数を評価するために、モデルの学習が必要となる。このため、HPO においては、多くの場合、目的関数の評価が実行時間におけるボトルネックとなり、評価回数と実行時間のトレードオフになる [4]。効率的にハイパーパラメータ空間を探索する様々な手法が提案されている。その手法の1つとして、自然界の生物の群れが高度な振る舞いをするをコンピュータに適用した群知能がある。郡知能には ABC, PSO, ACO などのアルゴリズムがある [abc][5][6]。これらのアルゴリズムを SVM のハイパーパラメータ最適化や特徴選択に適用した事例がある []。

郡知能の中でも ABC はパラメータが少なく比較的シンプルなアルゴリズムである。これらの研究では、SVM のカーネル関数を固定して実験を行っている。SVM には、様々なカーネル関数が適用でき、それぞれのカーネル関数はハイパーパラメータの数が異なる。ゆえにカーネル関数もハイパーパラメータとして扱うことができる。そこで、本研究では C に加えて 4 つのカーネル関数 (線形, RBF, シグモイド, 多項式) とそのハイパーパラメータも最適化対象とし、ABC を用いて SVM のハイパーパラメータ最適化を行う。

2 サポートベクトルマシン (SVM)

SVMは1995年にCortesらによって提案された機械学習アルゴリズムである[3]。SVMは教師あり学習を用いる2値分類器であるが、多クラスへの拡張も可能である。SVMは、訓練サンプル集合からデータを分類する識別関数を学習する。この学習過程において、SVMは訓練データの中で識別関数に近いデータであるサポートベクトルを得る。そして、サポートベクトルと識別関数との距離、すなわちマージンを最大化するように識別関数を構築する。これにより、SVMは新しいデータを分類する際、最大限のマージンを持ってデータを分類できるようになり、未知のデータに対しても誤分類のリスクを最小限に抑えることができる。

2.1 ハードマージン SVM

n 個の N 次元教師データ $\mathbf{x}_i (i = 1, \dots, n)$, がクラス K_1, K_2 の2値分類データであるとして、対応するクラスラベルをクラス K_1 の時に $y_i = 1$, クラス K_2 の時に $y_i = -1$ とすると、データが線形分離可能であれば識別関数は(2.1)式のように定義される。 \mathbf{w} は m 次元ベクトル、 b は識別関数のパラメータである。

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (2.1)$$

すべての教師データ $\mathbf{x}_i (i = 1, \dots, n)$ について(2.2)式の条件を満たすように \mathbf{w}, b を調節する事がSVMの学習フェーズとなる。

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}_i + b = \begin{cases} \geq 1 & \mathbf{x}_i \in K_1 \\ \leq -1 & \mathbf{x}_i \in K_2 \end{cases} \quad (2.2)$$

このとき、(2.2)式は(2.3)式と等価になる。

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad (2.3)$$

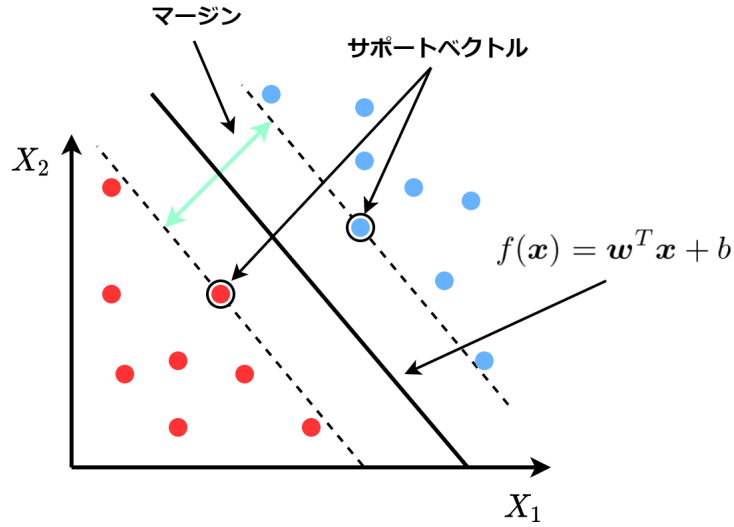


図 1: SVM の分類方法

ここで図 1 にデータを分離する超平面 (2 次元では直線) を示す. 分離超平面とそれに最も近いデータ (サポートベクトル) との距離をマージンと呼び, SVM はマージンを最大化するように分離超平面を求める. マージンを最大化することで, サポートベクトルから少しずれたデータが存在しても分類可能になり, モデルの汎化性能が向上する. 次に, マージンを最大化する超平面を求める方法を導く. \mathbf{x}_i から超平面への距離を d とすると, d は (2.4) 式のようなになる.

$$d = \frac{|f(\mathbf{x}_i)|}{\|\mathbf{w}\|} \quad (2.4)$$

よって, マージンを M とすると, マージン最大化は (2.5) 式の最適化問題によって表現できる

$$\max_{\mathbf{w}, b} M, \text{ subject to } \frac{y_i(\mathbf{w}^T \mathbf{x}_i + b)}{\|\mathbf{w}\|} \geq M \quad (i = 1, \dots, n) \quad (2.5)$$

さらに, (2.5) 式の制約条件部分を M で割り, $\frac{\mathbf{w}}{M\|\mathbf{w}\|}$ と $\frac{b}{M\|\mathbf{w}\|}$ をそれぞれ $\tilde{\mathbf{w}}, \tilde{b}$ とすると制約条件は (2.6) 式に書き換えられる.

$$y_i(\tilde{\mathbf{w}}^T \mathbf{x}_i + \tilde{b}) \geq 1 \quad (2.6)$$

(2.6) 式の等号が成立するデータ \mathbf{x}_i は分離超平面と距離が最も近いデータ, すなわちサポートベクトルである. したがってマージン \tilde{M} は (2.7) 式のようなになる

$$\tilde{M} = \frac{1}{\|\tilde{\mathbf{w}}\|} \quad (2.7)$$

よって (2.5) 式は (2.8) 式で表すことができる.

$$\max_{\tilde{\mathbf{w}}, \tilde{b}} \frac{1}{\|\tilde{\mathbf{w}}\|}, \quad \text{subject to } y_i(\tilde{\mathbf{w}}^T \mathbf{x}_i + \tilde{b}) \geq 1 \quad (i = 1, \dots, n) \quad (2.8)$$

$\|\tilde{\mathbf{w}}\|$ は正であるため, $\frac{1}{\|\tilde{\mathbf{w}}\|}$ を最大化する問題は, $\|\tilde{\mathbf{w}}\|^2$ を最小化する問題に等しい. $\tilde{\mathbf{w}}, \tilde{b}$ を改めて \mathbf{w}, b と表記すると (2.8) 式は (2.9) 式に書き換えられ, マージン最大化は (2.9) 式を解く問題に帰着できる.

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2, \quad \text{subject to } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad (i = 1, \dots, n) \quad (2.9)$$

さらに, 制約付き最適化問題である (2.9) 式をラグランジュの未定乗数法により, 双対問題に置き換える. ラグランジュ乗数 $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$ を用いてラグランジュ関数 L を (2.10) 式と定義する.

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{n=1}^n \alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1\} \quad (2.10)$$

制約条件が不等式であるため, \mathbf{w} の最適解において以下の KKT 条件が適用できる.

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = 0 \quad (2.11a)$$

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = 0 \quad (2.11b)$$

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 = 0 \quad (2.11c)$$

$$\alpha_i \geq 0 \quad (2.11d)$$

(2.11c) 式より, $\alpha_i = 0$ または $\alpha_i > 0$ で $y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$ が満たされなければならない. ここで, $\alpha_i > 0$ となる \mathbf{x}_i をサポートベクトルと呼ぶ. SVM は, 教師データ中から, サポート

トベクトルのみを用いて識別関数を構成する．(2.11a) 式， (2.11b) 式を計算すると，それぞれ (2.12a) 式， (2.12b) 式になる．

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{w}_i \quad (2.12a)$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (2.12b)$$

(2.12a) 式， (2.12b) 式を (2.10) 式に代入すると，(2.13a)，(2.13b) 式の， α のみで表されるハードマージンの双対問題が得られる．これは線形分離可能な場合のみに適用でき，誤分類を許容しない．

$$\max_{\alpha} \left\{ \tilde{L}(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\} \quad (2.13a)$$

$$\text{subject to } \sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0, (i = 1, \dots, n) \quad (2.13b)$$

ここで，サポートベクトルに対応する添字の集合を S とすると，(2.12a) 式より，識別関数は (2.14) 式になる．

$$D(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \quad (2.14)$$

また，(2.11c) 式より， b は (2.15) 式になる．

$$b = y_i - \mathbf{x}_i^T \mathbf{x} (i \in S) \quad (2.15)$$

(2.14)，(2.15) より得られた識別関数により， \mathbf{x} は $D(\mathbf{x}) > 0$ ならクラス K_1 に， $D(\mathbf{x}) < 0$ ならクラス K_2 に分類する．

2.2 ソフトマージン SVM

2.1 節で述べたハードマージン SVM は線形分離可能なことを前提としているが，現実の問題は非線形な問題が多い．そこで非線形な問題に対しても適応できるように，(2.3) 式に

非負の変数 $\xi_i (\geq 0)$ を導入する．その式を (2.16) 式に示す．

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i \quad (2.16)$$

これにより，マージンの内側にデータが存在することを許容する．このときソフトマージンの最適化問題は，(2.17a) 式，(2.17b) 式になる．ここで C は誤分類をどれだけ許容するかを決めるハイパーパラメータであり，小さいほど誤分類を許容し，大きいほど誤分類を許容しない．よってこの問題は，(2.17a) 式の第 1 項のマージン最大化と第 2 項の誤分類の許容数のバランスを図る問題である．

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (2.17a)$$

$$\text{subject to } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \ (i = 1, \dots, n) \quad (2.17b)$$

2.1 節と同様に，(2.17a) 式，(2.17b) 式を主問題として，ラグランジュの未定乗数法により双対問題を導出する．ラグランジュ乗数 $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$ ， $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)^T$ を用いてラグランジュ関数 L を (2.18) 式と定義する．

$$L(\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i\} - \sum_{i=1}^n \beta_i \xi_i \quad (2.18)$$

制約条件が不等式であるため， \mathbf{w} の最適解において以下の KKT 条件が適用できる．

$$\frac{\partial L(\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \mathbf{w}} = 0 \quad (2.19a)$$

$$\frac{\partial L(\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial b} = 0 \quad (2.19b)$$

$$\frac{\partial L(\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \xi} = 0 \quad (2.19c)$$

$$\alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i\} = 0 \quad (2.19d)$$

$$\beta_i \xi_i = 0 \quad (2.19e)$$

$$\alpha_i \geq 0, \beta_i \geq 0, \xi_i \geq 0 \quad (2.19f)$$

(2.19a)～(2.19c) 式を計算すると, それぞれ (2.20a)～(2.20c) 式になる.

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{w}_i \quad (2.20a)$$

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (2.20b)$$

$$\alpha_i + \beta_i = C \quad (2.20c)$$

(2.20a)～(2.20c) 式を (2.18) 式に代入すると, α のみで表されるソフトマージンの双対問題 (2.21a), (2.21b) 式が得られる.

$$\max_{\alpha} \left\{ \tilde{L}(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\} \quad (2.21a)$$

$$\text{subject to } \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, (i = 1, \dots, n) \quad (2.21b)$$

識別関数はハードマージン SVM と同じく (2.14), (2.15) 式になる. よってソフトマージン SVM とハードマージン SVM の違いは, α_i の上限 C の有無だけの違いである, 本研究ではハイパーパラメータである C を最適化対象に含める.

2.3 カーネルトリック

2.2 節で述べたソフトマージン SVM によって, 非線形な問題に対しても分類できるようになったが, 非線形で複雑な分類問題に対しては高性能な分類器を構成できない. そこで SVM は, N 次元特徴ベクトルを N' 次元特徴空間に写像し, 特徴空間上で分離超平面を得る. 高次元空間への写像の例を図 2 に示す. この例では, $X_1 X_2$ の軸を新たに追加することによって元の次元では非線形なデータを線形分離可能にしている. 一般に, 線形分離可能性はデータのサンプル数が多くなればなるほど難しくなり, 特徴空間の次元数が大きくなるほど易しくなる. 特に, $n-1$ 次元空間では, 最大で n 個のデータを任意のラベル付けで線形分離可能である. そのため N 次元特徴ベクトルを N' 次元特徴空間に写像することで

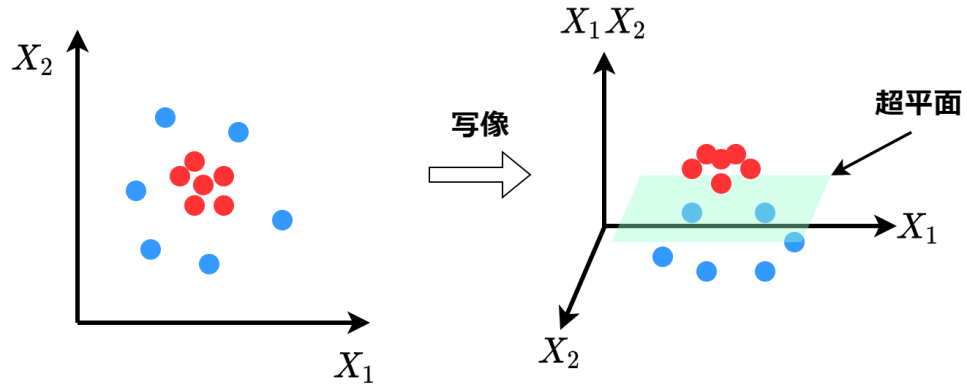


図 2: 高次元空間への写像の例

線形分離可能性が向上する．ここで， N 次元特徴ベクトルを N' 次元特徴空間に写像する関数を $\phi(\mathbf{x}) = \{\phi_1(\mathbf{x}), \dots, \phi_{N'}(\mathbf{x})\}^T$ とすると，特徴空間上での分離超平面は (2.22a)，(2.22b) 式の最適化問題を解くことで得られる．

$$\max_{\alpha} \left\{ \tilde{L}(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \right\} \quad (2.22a)$$

$$\text{subject to } \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, (i = 1, \dots, n) \quad (2.22b)$$

しかし次元 N' が大きくなればなるほど $\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ の計算量が膨大になる．そこでカーネル関数を (2.23) 式で定義する．

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (2.23)$$

もし，(2.23) 式を満たすようなカーネル関数が存在するならば，入力特徴ベクトルの内積計算で $\phi(\mathbf{x})$ を得ることができる．すなわち計算量が膨大になる可能性がある $\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ を直接計算する必要がない．このように内積計算をカーネル関数で置き換える手法をカーネルトリックと呼ぶ．本実験では，(2.24)～(2.27) 式で表されるカーネル関数をハイパーパラメータの対象とする．また，(2.24)～(2.27) 式内にあるハイパーパラメータは γ , coef0 , d

の3つである.

$$\text{線形カーネル: } K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \cdot \mathbf{x}_j \quad (2.24)$$

$$\text{RBF カーネル: } K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (2.25)$$

$$\text{シグモイドカーネル: } K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \cdot \mathbf{x}_j + \text{coef0}) \quad (2.26)$$

$$\text{多項式カーネル: } K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \cdot \mathbf{x}_j + \text{coef0})^d \quad (2.27)$$

カーネル関数の条件を挙げるべきか (勉強不足)

2.4 多クラス分類への拡張

SVMは2値分類器であるが, 1対他法や1対1法により多クラス分類への拡張ができる. 1対他法では, 各クラスに対し, そのクラスとそれ以外のクラスの分類をする. そのためクラスが K 個ある場合, K 個の SVM モデルを用意する必要がある. 一方, 1対1法では各クラスのペアの組み合わせに対して分類をする. そのためクラスが K 個ある場合, $\frac{K(K-1)}{2}$ 個の SVM モデルを用意する必要がある. 以上より, より計算量が少ないのが1対他法, より詳細な分類をするのが1対1法の特徴である. 本研究では一対他法を使用する.

3 ABC アルゴリズム

3.1 概要

郡知能とは個々での単純な行動が、集団としては高度な振る舞いをすることを模した最適化アルゴリズムである。ABC は 2005 年に karabara によって提案された郡知能の一つである [4]。ABC は蜜蜂の採餌行動から着想を得ており、収穫蜂、追従蜂、偵察蜂の 3 種類の蜂によって探索を行う。収穫蜂はすべての食物源に対して探索を行い、追従蜂は評価値の高い食物源を優先して探索を行なう。追従蜂の探索によって評価値の高い食物源の近傍を局所的に探索することができる。収穫蜂と追従蜂による探索で改善されなかった食物源は偵察蜂によってランダムな新しい食物源に置き換えられる。偵察蜂によって局所最適解からの脱出が可能になり、探索範囲を広げることができる。これら 3 種類の蜂による探索を繰り返すことで、より良い解を求めていく。

3.2 探索手順

ABC によって、(3.1) 式のような目的関数を $f(\mathbf{x})$ を最小とする D 次元の \mathbf{x} を求める問題を解くことを考える。

$$\min. f(\mathbf{x}), \text{ subject to } \mathbf{x} \in \mathbb{R}^D \quad (3.1)$$

ABC では食物源の数 n と探索上限回数 $limit$ の 2 つのパラメータをあらかじめ設定する必要がある。食物源の数を n とすると、収穫蜂、追従蜂の数はそれぞれ n ずつとなる。 $limit$ は偵察バチフェーズで使用する食物源の探索回数の上限である。この値を小さくするとより広範囲の探索に、大きくするとより局所的な探索を行う。各食物源に番号を振り分け、 i 番目の食物源を $\mathbf{x}_i (i = 1, \dots, n)$ とすると \mathbf{x}_i の評価値 $fit(\mathbf{x}_i)$ は (3.2) 式の評価関数により求

められる.

$$fit(\mathbf{x}_i) = \begin{cases} \frac{1}{1 + f(\mathbf{x}_i)} & \text{if } f(\mathbf{x}_i) \geq 0 \\ |1 + f(\mathbf{x}_i)| & \text{otherwise} \end{cases} \quad (3.2)$$

step 0 準備

ABC におけるパラメータである食物源の数 n と探索上限回数 $limit$ の設定を行う. 次に終了条件の設定を行なう. 終了条件としては, アルゴリズムの反復回数が最大反復回数を超えたとき, 最良の食物源の評価値もしくは目的関数値が一定値を超えたときなどがある.

step 1 初期化

初期化では, (3.3) 式によって各食物源を各成分の範囲内でランダムに生成する. さらに各食物源の探索回数 $trial_i$ を 0 に初期化する. x_{ij} は \mathbf{x}_i の j ($j = 1, \dots, D$) 番目の成分, $xmin_j$, $xmax_j$ はそれぞれ食物源の各成分の最小値, 最大値である. また, $rand$ は一様乱数を表す.

$$x_{ij} = xmin_j + rand[0, 1](xmax_j - xmin_j) \quad (3.3)$$

初期化された食物源の中で最も評価値が高いものを \mathbf{x}_{best} とし, その食物源のインデックスを i_{best} とする

step 2 収獲蜂フェーズ

収獲蜂フェーズでは収獲蜂が食物源の近傍を探索する. このときの更新式は (4.5) 式のようにになる. ここで j, k ($k \neq i$) はランダムに選択される.

$$v_{ij} = x_{ij} + rand[-1, 1](x_{ij} - x_{kj}) \quad (3.4)$$

ここで v_{ij} と x_{ij} の評価値を比較し, v_{ij} が x_{ij} よりも優れていたら x_{ij} を v_{ij} に置き換え, $trial_i$ を 0 にリセットする. そうでなければ, $trial_i$ を 1 増やす.

step 3 追従蜂フェーズ

追従蜂フェーズでは収獲蜂フェーズによって更新された食物源の評価値に基づいて、ルーレット選択を行い更新個体を選択する。食物源 x_i の選択確率 p_i は (4.6) 式のようになる。そのため、評価値が高い食物源ほど選択確率が高く、評価値が低い食物源は選択確率が低くなる。食物源の更新は収獲蜂フェーズと同様に行われる。この操作を n 回行う。

$$p_i = \frac{fit(\mathbf{x}_i)}{\sum_n fit_j} \quad (3.5)$$

step 4 偵察蜂フェーズ

偵察蜂フェーズでは、 $trial$ の値が探索打ち切り回数 $limit$ に達した食物源を (3.3) 式によりランダムに生成した新たな解に置換する。置換した食物源の $trial$ を 0 にリセットする。 $trial$ の値が $limit$ に達した食物源がなければ何も行わない。

step 5 終了判定

各食物源の評価値 $fit(\mathbf{x}_i)$ と、 $fit(\mathbf{x}_{best})$ を比較し、 $fit(\mathbf{x}_i) > fit(\mathbf{x}_{best})$ となる \mathbf{x}_i が存在する場合は \mathbf{x}_{best} を更新する。その後、終了条件を満たしている場合は最適解をその時の \mathbf{x}_{best} として探索を終了し、そうでない場合は **step 2** に戻り、探索を続ける。

4 提案手法

SVM のハイパーパラメータ最適化では、カーネル関数の選択も考慮する余地があるが、カーネル関数を固定して行われた研究が多い。本研究では、SVM のハイパーパラメータ C に加え、(2.24)～(2.27) 式の 4 つのカーネル関数とそのハイパーパラメータ $(\gamma, \text{coef0}, d)$ を最適化対象とする手法を提案する。最適化アルゴリズムは設定パラメータの少ない ABC を使用する。これにより、ハイパーパラメータの探索範囲を広げ、より良い SVM モデルの探索が可能になる。

4.1 ABC アルゴリズムの適用

(2.24)～(2.27) 式の 4 つのカーネル関数は、ハイパーパラメータの数や性質が異なっている。そのためカーネル関数が異なる個体では個体表現が変わり、個体の次元数が変化する。しかし、(2.25)～(2.27) 式の γ や、(2.26), (2.27) 式の coef0 はそれぞれスケール項、定数項と共通の性質を持っている。そこで、本研究では ABC の個体表現を (4.1) の 5 次元とし、個体表現の変化をその個体の各次元の活性、非活性により表現することとした。ここで、カーネル関数、 C は常に使用されるハイパーパラメータで $\gamma, \text{coef0}, d$ はその個体のカーネル関数によって活性、非活性になる。

$$\text{個体表現: } (C, \gamma, \text{coef0}, d) \quad (4.1)$$

カーネル関数の値は 4 つのカーネル関数を整数値にエンコードしたもので、 $C, \gamma, \text{coef0}, d$ は $[0.1]$ の範囲の実数とした。得られた実数を a とするとそれぞれのパラメータは (4.2) 式によって指定した範囲内で SVM に適用する値 A となる [2]。ここで a_{max}, a_{min} はパラメータの上限値、下限値を表している。

$$A = a(a_{max} - a_{min}) + a_{min} \quad (4.2)$$

表 1: カーネル関数によるパラメータの活性, 非活性

カーネル関数	γ	coef0	d
線形カーネル	0	0	0
RBF カーネル	1	0	0
シグモイドカーネル	1	1	0
多項式カーネル	1	1	1

また, d は離散値であるため, A を四捨五入したものを SVM に適用し, 非活性の次元は探索の際に選択されない. 表 1 にカーネル関数による活性, 非活性の状態の組み合わせを示す. ここで 1 が活性, 0 が非活性を表す.

また, ABC による最適化は連続値を前提としているため, カテゴリ変数であるカーネル関数に ABC の更新式は適用できない. そこで本研究では更新次元にカーネル関数が選択された際の更新は (4.3) 式の確率 P で更新されるものとする. ここで \mathbf{x}_i は更新個体, \mathbf{x}_j は \mathbf{x}_i と異なるカーネル関数を持つ個体の中からランダムに選ばれた個体である.

$$P = \frac{f(\mathbf{x}_j)}{f(\mathbf{x}_i) + f(\mathbf{x}_j)} \quad (4.3)$$

ABC における食物源の評価値 fit_i は (4.4) 式によって計算される. ここで $miss_i$ は SVM が検証セットの分類を行い, 誤分類した割合である.

$$fit_i = \frac{1}{1 + miss_i} \quad (4.4)$$

4.2 提案手法のアルゴリズム

提案手法のアルゴリズムを以下に示す. ただし評価値の計算は次の手順で行われる.

1. カーネル関数以外の値を (4.2) 式でマッピング.

2. その食物源が多項式カーネルだった場合 d の値を四捨五入.
3. 学習セットで svm の学習を行う.
4. 検証セットで svm の分類精度を算出.
5. (4.4) 式で評価値の計算

step 0 準備

ABC におけるパラメータである食物源の数 n と探索上限回数 $limit$ の設定を行う. 次に SVN のハイパーパラメータ C , γ , $coef0$, d の探索範囲の設定を行い, 4つのカーネル関数を整数値にエンコードする. 最後に終了条件であるアルゴリズムのサイクル数を設定する.

step 1 初期化

初期化では, n 個の食物源をカーネル関数を定める 1次元目を 1~4 の整数値でランダムに初期化, その他の 2~5次元は $[0,1]$ の一様乱数で初期化する. n 個の食物源の評価値を計算し, 初期化された食物源の中で最も評価値が高い食物源とそのインデックスを保存する.

step 2 収獲蜂フェーズ

収獲蜂フェーズでは収獲蜂が食物源の近傍を探索する. このときの更新式は (4.5) 式のようになる. ここで $j, k(k \neq i)$ はランダムに選択される.

$$v_{ij} = x_{ij} + rand[-1, 1](x_{ij} - x_{kj}) \quad (4.5)$$

ここで v_{ij} と x_{ij} の評価値を比較し, v_{ij} が x_{ij} よりも優れていたら x_{ij} を v_{ij} に置き換え, $trial_i$ を 0 にリセットする. そうでなければ, $trial_i$ を 1 増やす.

step 3 追従蜂フェーズ

追従蜂フェーズでは収獲蜂フェーズによって更新された食物源の評価値に基づいて、ルーレット選択を行い更新個体を選択する。食物源 x_i の選択確率 p_i は (4.6) 式のようになる。そのため、評価値が高い食物源ほど選択確率が高く、評価値が低い食物源は選択確率が低くなる。食物源の更新は収獲蜂フェーズと同様に行われる。この操作を n 回行う。

$$p_i = \frac{fit(\mathbf{x}_i)}{\sum_n fit_j} \quad (4.6)$$

step 4 偵察蜂フェーズ

偵察蜂フェーズでは、 $trial$ の値が探索打ち切り回数 $limit$ に達した食物源を step 1 の初期化と同じ手順でランダムな新たな解に置換する。置換した食物源の $trial$ を 0 にリセットする。 $trial$ の値が $limit$ に達した食物源がなければ何も行わない。

step 5 終了判定

各食物源の評価値と、この時点で最良の食物源の評価値を比較し、 $fit(\mathbf{x}_i) > fit(\mathbf{x}_{best})$ となる \mathbf{x}_i が存在する場合は \mathbf{x}_{best} を更新する。その後、終了条件を満たしている場合は最適解をその時の \mathbf{x}_{best} として探索を終了し、そうでない場合は step 2 に戻り、探索を続ける。

表 2: KDD'99 データセットの内訳

クラス	KDD'99	10%KDD'99
normal	972780	97279
dos	3883370	391458
probe	41102	4107
r2l	1126	1126
u2r	52	52
合計	4898430	494021

表 3: 実験データセットの内訳

クラス	学習	検証	テスト
normal	9740	9650	9766
dos	39127	39238	39106
probe	412	385	410
r2l	120	125	115
u2r	3	4	5
合計	49402	49402	49402

5 実験

侵入検知問題である 10%KDD'99 データセットを使用して, scikit-learn のデフォルトパラメータ, 既存手法, 提案手法の間で表 ?? の設定で比較実験を行った. 実験は, データセットからランダムに 10% 抽出した学習セット, 検証セット, テストセットの 3 つのデータセットを用意して行った. 検証セットは ABC アルゴリズムにおける解の評価に使われ, 得られた最適解を未知のデータであるテストセットの分類に適用し, その結果を最終的な分類精度とする.

5.1 データセット

10%KDD'99 データセットは, オリジナルの KDD'99 データセットのうちサンプル数の多い normal, dos, probe クラスを 10% に減らしたデータセットである. また, KDD'99 データセットの特徴数は 41 個である. KDD'99 データセットの内訳を表 2 に示す. 本研究では, ABC で得られた SVM モデルが未知のデータに対して有効であるかを評価するために 3 つのデータセットを用意した [2]. 3 つのデータセットは, SVM の学習に使用する学習セット,

表 4: SVC クラスのデフォルト設定

パラメータ	デフォルト値
kernel	rbf
C	1.0
γ	1/n_feature
coef0	0
degree(d)	3

SVM の評価に使用する検証セット，最適解を評価するためのテストセットである．ここでテストセットは ABC で得られた最適解を評価するために一度だけ使用される．3つのデータセットを内訳を表 3 に示す．これらのデータセットは 10%KDD'99 データセットからランダムに 10%抽出している．

5.2 実験設定

5.2.1 実験環境

SVM の実装には，CPU: Intel Core i7-12700，メモリ 32GB の計算機上で Python 3.11.0 と scikit-learn 1.5.0 ライブラリの SVC クラスを使用した．SVC クラスにおける本研究で使用するハイパーパラメータのデフォルト設定を表 4 に示す．ここで n_feature はデータセットの特徴数を表す．

5.2.2 パラメータ設定

表 5, 6 に本研究のパラメータ設定を示す

表 5: ABC の実験パラメータ

コロニーサイズ	20
LIMIT	100
サイクル数	500

表 6: SVM の実験パラメータ

kernel	[linear, rbf, sigmoid, poly]
C	$[10^{-6}, 35000]$
γ	$[10^{-6}, 32]$
coef0	$[0, 10]$
degree(d)	$[1, 3]$

5.3 実験結果

実験結果を表 7 に示す.

表 7: 実験結果

	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	99.84	99.90
実行時間 [h]	-	-	-	-	11.7	7.0

6 考察

内容 1

7 おわりに

内容 1

謝辞

山口先生有り難う！

参考文献

- [1] M. Feurer and F. Hutter, “Hyperparameter optimization,” Automated Machine Learning, pp.3–33, Springer, 2019.
- [2] 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.
- [3] Cortes, C. and Vapnik, V. Support-vector networks, Machine Learning, Vol.20, No.3, pp.273-297, 1995.
- [4] 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.