

ABCにおけるハイパーパラメータの複合最適化法の提案と評価

日大生産工 ○松井 翔吾 日大生産工 山内 ゆかり

1. まえがき

近年のディープラーニングやビッグデータの活用の急速な進展に伴い、機械学習全般への注目が高まり、国内外において活発に研究が行われている。機械学習を用いて分類問題や回帰問題に対して優れた学習モデルを得るために、各学習手法のハイパーパラメータの最適化やデータセットからの特徴選択が重要となり、それらを得るための代表的な学習手法としてランダムフォレスト (Random Forests: RF)[1]とサポートベクトルマシン(Support Vector Machines: SVM)[2]がある。

RFには予め調整が必要なハイパーパラメータが存在し、SVMはより高精度な分類(回帰)を可能にするためにハイパーパラメータを調整し、特徴選択を行う必要がある。そこで、ハイパーパラメータの最適化やデータセットから特徴選択を行う手法として群知能 (Swarm Intelligence: SI)[3]の1つの人工蜂コロニー (Artificial Bee Colony: ABC)アルゴリズムが注目されている。[4]

近藤らはハイパーパラメータの調整の負担の軽減と分類精度の優れた学習モデルの取得をするために、SIよりも実行時間を必要としないABCによる、RFとSVMのハイパーパラメータの最適化とデータセットからの特徴選択を同時に行うことを提案した。しかし、不要な特徴を多く削減できていなくても、分類精度が高ければ、暫定的最適解として保存され続ける問題が生じる。

本研究では、ABCの修正を行うことで、不要な特徴を多く削減できていなくても、分類精度の高さに影響されずに、暫定的最適解に保存され続けずに改善することを目的に報告する。

2. 従来手法

2-1. ABC

ABCは2005年にD.Karabogaによって提案された群知能の1つであり、収獲蜂、追従蜂、偵察蜂の3種類の蜜蜂の採餌行動に基づいた3フェーズからなる。収獲蜂と追従蜂のフェーズでは、解の候補近傍の局所探索を行うが、収獲蜂は決定論的に解の候補を選択し、追従蜂は蓋然的に

解の候補を選択する。偵察蜂のフェーズは、採餌行動において尽きた餌場を消去する行動を基にしたもので、探索空間の新たな領域を探索するために、新たな解を挿入する。収獲蜂の数を n とすると追従蜂の数も n であり、巣を構成する蜜蜂の総数(コロニーサイズ)は $2n$ となる。餌場(候補解)の数は収獲蜂と同数の n である。ABCの枠組みをFig.1に示す。

```
while 停止条件を満たさない do
  収獲蜂フェーズ: 各候補解の更新を行う.
  追従蜂フェーズ: 各候補解の適応度に基づいて
                  確率によって候補解を選択し、
                  その候補解の更新を行う.
  偵察蜂フェーズ: 設定された回数の間に1度も更新
                  されなかった候補解の1つを
                  ランダムに生成し新たな候補解
                  と置換する.
end
```

Fig.1 ABCの枠組み

2-2. 収獲蜂フェーズ

候補解 x_i を用いて下記の更新式(1)により更新(ここでは $v_i = x_i$)を行う。

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{jk}) \quad (1)$$

ここで j はランダムに選択される。 $x_k(k \neq i)$ はランダムに選択された x_i 以外の候補解となり、 ϕ_{ij} は $[-1,1]$ 中の一様乱数である。 v_i と x_i の適応度を比較し、 v_i が優れていれば x_i と置換し、そうでない場合は x_i の試行カウンタを1増やす。これを $i = 1, \dots, n$ に対して繰り返す。対象問題が目的関数 $f(x)$ の値を最小とする x の値を求める最小化問題ならば、適応度は下記の評価関数(2)によって求められる。

$$fit_i = \begin{cases} \frac{1}{1 + f(x_i)} & \text{if } f(x_i) \geq 0 \\ 1 + \text{abs}(f(x_i)) & \text{if } f(x_i) < 0 \end{cases} \quad (2)$$

2-3. 追従蜂フェーズ

収獲蜂フェーズで更新された候補解の適応度 fit_i に基づいた確率 $p_i = fit_i / \sum_n fit_i$ を用いて更新対象となる候補解を選択し、その候補解に対して収獲フェーズと同様の更新操作を n 回繰り返す。適応度に基づく確率によって更新対象の

Proposal and Evaluation of Compound Optimization Method
for Hyperparameters by ABC

Shogo MATSUI and Yukari YAMAUCHI

選択が行われるため、適応度の高い候補解は繰り返し更新操作の対象になる。

2-4. 偵察蜂フェーズ

収穫フェーズと追従フェーズを通して予め決められた打ち切り回数(*limit*)の間に1度も更新されなかった解をランダムに生成した解に置換する。この操作は1サイクルに1回程度しか行われず、コロニーに多様性がもたらされ、局所解に陥ることを防ぐことができる。ABCアルゴリズムによる最適化手順をFig.2に示す。

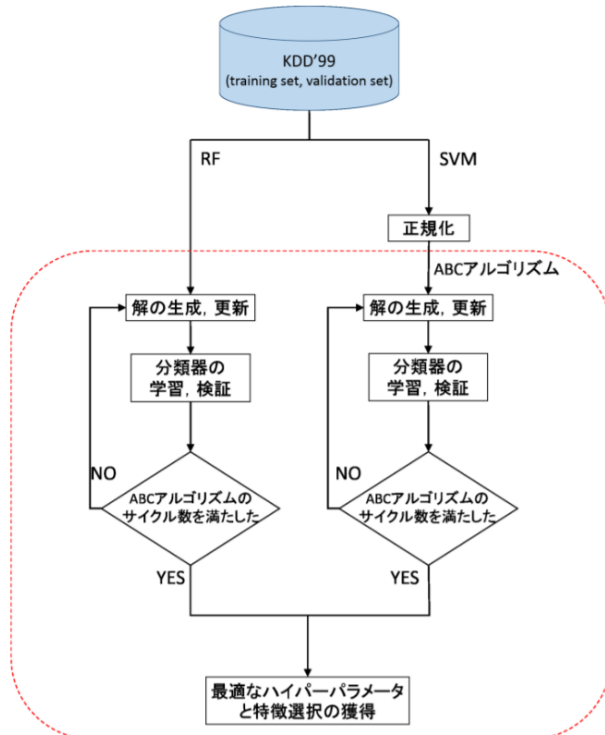


Fig.2 ABC アルゴリズムによる最適化手順

3. 提案手法

従来手法におけるパラメータの最適化とデータセットからの特徴選択を同時に行う際に生じる、不要な特徴を多く削減できていなくても、分類精度の高さに影響されずに、暫定的最適解に保存され続けるという問題の修正を行い、精度の向上を図る。

4. 実験

本実験では群知能の1つであるABCを実装し、従来手法の実験で扱っている侵入検知問題のベンチマークとして広く利用されているKDD'99 Cup 1999 data (KDD'99)を用い、分類精度、検知率(攻撃を攻撃として検知できた割合)および誤警報率(正常状態を攻撃として検知した割合)における従来手法と提案手法の比較を行うことで問題の解決を試みる。

5. まとめ

本研究ではABCによるRFとSVMのハイパーパラメータの最適化とデータセットからの特徴選択を同時に行う際に生じる、不要な特徴を多く削減できていなくても、分類精度が高ければ、暫定的最適解として保存され続けるという問題に対し、ABCの修正を行うことで、不要な特徴を多く削減できていなくても、分類精度の高さに影響されずに、暫定的最適解に保存され続けない様に改善することを試みた。現段階では実験に着手することができていないため、この研究にはまだ取り組むべきことが多くある。

参考文献

- [1] L.Breiman, "Random Forests", Kluwer Academic Publishers. Manufactured in The Netherlands., Machine Learning, 45, 5-32, (2001)
- [2] C.Corters and V.Vapnik, "Support-Vector Networks", Kluwer Academic Publishers, Boston. Manufactured in The Netherlands., Machine Learning, 20, 273-297, (1995)
- [3] G.Beni, "Swarm Intelligence" Springer Science+Business Media New York, (2014)
- [4] 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の最適化と特徴選択」原著論文