

Boosting Algorithm の示す複雑適応性について

An Investigation of Boosting Algorithm as Complex Adaptive System

三宅雅矩*¹
Masanori Miyake

池上高志*¹
Takashi Ikegami

*¹ 東京大学大学院総合文化研究科
Graduate School of Arts and Sciences, University of Tokyo

The ensemble learning in a machine learning field such as Swarm Optimization and AdaBoost have been paid much attention to. In particular, AdaBoost is a notable algorithm, which achieves an intelligent system based on an ensemble of “weak” learners. AdaBoost shows excellent performance in both classification and regression tasks. Such ensemble learning systems have been also studied as complex adaptive systems, comparing with natural immune networks or traditional classifiers. A purpose of this paper is to examine the similarities of ensemble learning systems, and to construct integrated viewpoints. We will propose an extended AdaBoost algorithm to couple with a genetic algorithm, and evaluate its property as a complex adaptive system.

1. はじめに

近年、複数の学習器を統合し合議を形成することで、精度の高い学習器を構築する、アンサンブル学習とよばれる手法が統計的機械学習の分野で盛んに研究されている。その中でも特に AdaBoost[Freund and Schapire, 1997]のような、弱学習器の集合によって強学習器を生成する Boosting とよばれる手法が回帰問題・分類問題ともに高い性能を示し、有効であることが明らかになっている。一方、Learning Classifier System[Booker et al. 1989]や(人工)免疫システム、また粒子群最適化法を用いたシステムなども多数の行動主体を構成要素とし、これらの集団が自律分散的にシステム全体の意思決定を行うという点では前述のアンサンブル学習と類似した機構を有しているように思われる。

そこで、本研究ではこれらのシステムについて類似性を調査・吟味し、アンサンブル学習と複雑適応系の統一的視座を構築することを目的として AdaBoost のアルゴリズムにいくつかの改良を加えた実験を行った。結果、弱学習器のモデル選択および新奇性の導入といった点で多少の変更を行うことで、AdaBoost はその統計的学習手法としての有効性を失うことなく、複雑適応系の枠組みへの適用が可能であることを検証・確認する。

2. モデル

2.1 AdaBoost

Boosting 手法の一つである AdaBoost は、複数の弱学習器を逐次付け足していくことで強学習器を構築するが、このとき各弱学習器は正確性に応じた重み付けがなされ、正確なものほど全体に対し高い影響力を持つことになる。また、弱学習器が付け加えられる度に、その時点でうまく分類・回帰できなかったデータが次のステップで高い影響力を持つように、サンプルデータの重み付けが更新される。

AdaBoost においては分類問題・回帰問題ともに、しばしば弱学習器として決定木が用いられる。図 1 は、決定木および AdaBoost による回帰を行った例である。緑線と黄線はそれぞれ

深さ 1、深さ 2 の決定木が表現する回帰曲線の例であり、このような決定木を 100 個用いて構成された AdaBoost の強学習器による回帰曲線が、それぞれ赤線と青線に対応する。この例では、決定木の深さが 1 から 2 へと増すことで、より複雑な特徴表現の学習・抽出が可能となっている。

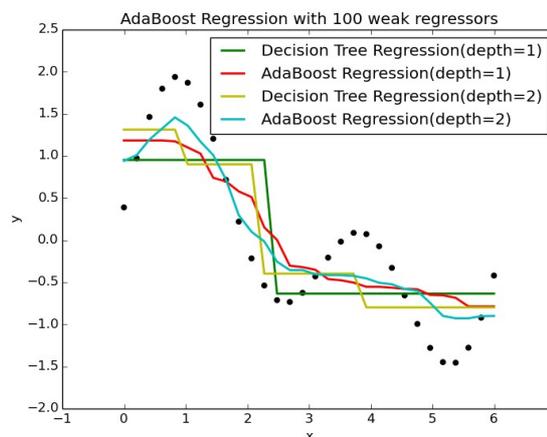


図 1: 深さ 1 の決定木による回帰曲線の例を緑色、深さ 2 の決定木による回帰曲線の例を黄色で示している。これらは、それぞれランダムに複数生成した中から最もデータへの当てはまりが良いものを選択している。また、このような深さ 1, 2 の決定木を弱学習器として用いた AdaBoost による回帰曲線を、それぞれ赤線および青線で示す。ともに 100 個の決定木を弱学習器として用いている。

AdaBoost において決定木を弱学習器として用いる場合、通常は使用する木の深さの最大値だけをモデルの複雑さの上限として最初に与えることが多い。複雑なモデルを表現できるようにするためには、この木の深さの最大値にある程度大きな値を設定することが必要とされる。このようにして設定された決定木の深さについて、余分な枝を後から剪定することで木の深さを適切な大きさに保とうとすることはあるが[Margineantu, 1997]、学習の状況に応じて付加する木の深さを AdaBoost のアルゴリズムが決定し、逐次変化させていくといった試みは今までなされて

いなかった。これはすなわち、AdaBoostにおける弱学習器のモデルが不要な複雑さを持たないように適応的に調整・決定する必要性については議論されてこなかったことを意味する。この点について、後述する他の複雑適応系との比較および改良方法の提案・実験の考察において議論・検討を行う。

なお、Boostingにおいては弱学習器が逐次的に付加されていくのは前述した通りであるが、並列処理を可能とするアルゴリズムは既に提案されている[Lazarevic et al., 2002]。また、通常Boostingはバッチ学習において有効であるが、オンライン学習においても成功を収めている[Oza, 2005]。

2.2 Learning Classifier System

Learning Classifier Systemは内部にClassifierとよばれるBit Stringで表現された条件と行動のルールセットを有する学習機構であり、この複数のClassifierが環境との報酬を介した相互作用を行うことで学習が進行する。各Classifierは環境だけでなく他のClassifierに対しても働きかけることができ、内部で報酬が伝播するネットワークを構築する。また、Bit Stringで表現される各Classifierの条件部には、0と1の他に#(Don't Care)とよばれるワイルドカードとなる文字が含まれており、これによりルールの一般化を行っている。システム内部のClassifierには蓄積された報酬(強度とよばれる評価値で表される)に応じてGAが適用され、環境に適合しない個体は淘汰される。

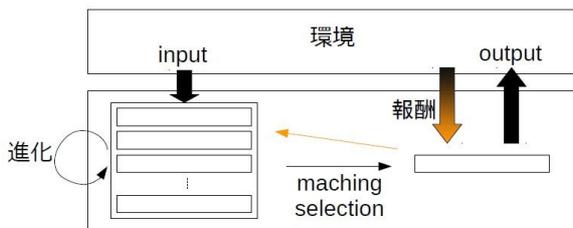


図 2: 環境の状態と適合する条件部を有するClassifierがmatching selectionによって選択され、環境に対して働きかけ、報酬を得ることができるようになる。複数のClassifierが選択された場合は強度が高く、#(Don't Care)の数が少ない(特異性の高い)ものが環境に対して働きかける権利を得る。

Learning Classifier System内の各Classifierは、環境の状態に応じて報酬を受け取るだけであり、明示的にパフォーマンスの評価値が与えられるわけではない。また、常に状態が変化するような環境に対して適用される。環境をうまく設定することで分類・回帰問題を扱うことはできるが、以上のような点で通常の教師あり学習とは異なり、また統計的な解析も困難である。このため、統計的機械学習の枠組みでAdaBoost等のアンサンブル学習と比較されることは少なく、マルチエージェントシステムといったような複雑適応系としての枠組みの中でのみ議論されることが多い。

しかしながら、両者には共通する点も多く存在する。次節では前述の議論を踏まえ両者を対比させることで、統一的視座の構築を試みる。

3. モデルの比較

AdaBoostとLearning Classifier Systemの両者に共通する最大の特徴は、複数の構成要素の集合としてシステムが構成されている点であった。このそれぞれの構成要素が有する性質・特徴を表1に整理する。

	Classifier System	AdaBoost
領域	複雑適応系システム	統計的機械学習
多数の構成要素	Bit String	弱学習器
ネットワーク性	Bit String間での報酬の伝播	学習器の逐次付加および重み付け
分散性	各Bit Stringが環境を知覚・行動	各弱学習器がデータを認識・処理
多様性	2^l の表現が可能(l : Bit String長)	弱学習器のモデルに依存
内部モデル	条件に対応づけられた行動	分類・回帰
処理方法	並列処理 オンライン処理	逐次処理 バッチ処理
学習機構	環境からの報酬により適応度を更新	正確性を基にした教師有り学習
新奇性の導入	GA	-

表 1: AdaBoostとLearning Classifier Systemの比較

ここで挙げた性質のうち、多数の構成要素・ネットワーク性(非線形性)・分散性・多様性・内部モデルについては複雑適応系の有する性質としてしばしば言及されるものである。例えば(人工)免疫システムであれば 10^8 を超える種類の免疫細胞が構成要素となり、それぞれがイデオタイプ・ネットワークを構成しながら認識可能な抗原に対して自律分散的に作用する。また、粒子群最適化を行うような群知能においては、実数値で表現される位置と速度を有した多数の粒子が構成要素となり、周囲の粒子と情報の伝達を行いながらやはり自律分散的に解空間を探索する。

Classifier System, AdaBoostについても、表1に示したようにともに対応する性質が存在し、その類似性が見て取れる。多様性についてはAdaBoostにおいては弱学習器の生成モデルに依存するものの、任意の値を取ることができる場合が多く、やはり非常に多様な表現が可能である。

処理方法については、Classifier Systemでは並列にオンライン処理を行っているのに対し、AdaBoostでは基本的に逐次的にバッチ処理を行うといった相違が存在する。しかし前述のように、並列・オンライン処理を可能とするAdaBoostアルゴリズムが近年考案され進展していることから、ここでも両者の対応関係が構成されている。

しかしながら、特徴選択における新奇性の導入に関しては、Classifier SystemにおけるGAのような機構がAdaBoostには存在しない。これは前述したような、学習の状態に応じて弱学習器のモデルを適応的に調整する機構がAdaBoostには存在しないことと関係しており、モデルの選択・評価機構とともに導入することが、複雑適応系としての理解のためには必要であると考えられる。

そこで、次節ではAdaBoostで使用する弱学習器のモデルを学習の進度に応じて変化させるような機構を導入することを提案し、この手法の有効性について実験およびその結果とともに議論を行う。

4. 実験

4.1 問題設定および手法

今回は AdaBoost.R2 を用いて回帰問題を扱うことを考える [Drucker, 1997]。損失関数として指数損失関数を、データセットとして UCI Machine Learning Repository の Geographical Origin of Music Data Set [Zhou et al., 2014] を用いた (なお、目的変数としては 2 つある中の longitude を選択した)。

AdaBoost の弱学習器としては、決定木を使用する。この決定木は毎回ランダムに複数生成した中から最も良いものを選択して付加する。決定木の深さは AdaBoost の学習が開始した時点では 1 であるが、そのモデルを用いた学習がある程度収束した時点で、順次決定木の深さを増やしていく、といった方針を取る。

回帰曲線の当てはまり具合については、決定係数を用いて評価する。また今回、学習がある程度収束にした時点については、その評価規準として

$$\begin{aligned} & \text{(直近 7 ステップ間での決定係数の変動)} \\ & < 0.004 / \text{(現在の決定係数)} \end{aligned}$$

を用いた。

4.2 結果および考察

前述のデータセットを適用した結果を、図 3 に示す。

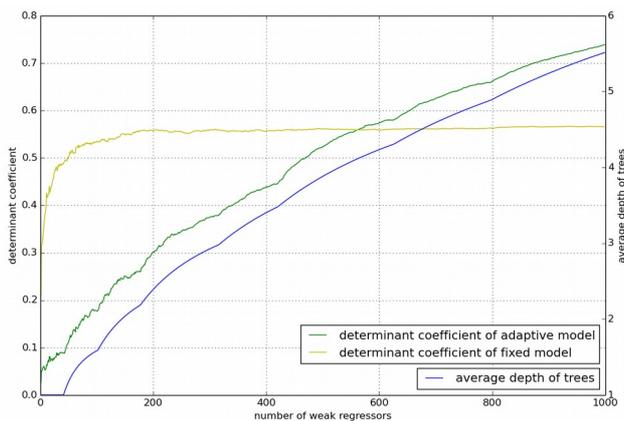


図 3: 黄線は深さが 5 で固定された決定木を AdaBoost の弱学習器として用いた場合の決定係数の変化、緑線は今回の提案手法を用いた場合の決定係数の変化である。また、青線は今回の提案手法を用いた場合の平均の決定木の深さである。黄線は初期の段階で急激に学習が進むが、その後は精度は改善せず、最終的には提案手法の方が当てはまり具合が良くなっていることが見て取れる。

この結果より、従来の弱学習器のモデルでは捉えきれなかった特徴を、使用する弱学習器のモデルを適応的に変化させ、新奇性を取り入れることで捉える能力を AdaBoost が有していることが確認できた。またこれは、複雑度が未知のデータに対しても、AdaBoost が学習を行いながら弱学習器のモデルを変化させることで適応できることを意味している。

今回はバッチ学習についてのみ扱ったため、決定木の深さを増やしていく方向へのみ拡張することを考えたが、オンライン学習について今回の手法を適用することが可能となれば、決定木の深さを減らし、不要なモデルの複雑さを取り除く方向への拡張も可能であると考えられる。また、決定木の深さが増した場合

に、既に加えた弱学習器を GA などと組み合わせることで、より表現を拡張することが可能となるのではないかと考えた議論も考えられる。

また今回、新奇性を導入するタイミングとしてある程度学習が収束した点を用いたが、このタイミングについてもより適切な選択方法を一般化した形で考える必要がある。これらの点に関して、当日は議論を行いたい。

5. 結論

本研究では AdaBoost と複雑適応系の学習システムとの比較・検討を行い、また AdaBoost を複雑適応系の学習システムとして捉える上で、適切な弱学習器のモデル選択を行い、新奇性を導入する機構をいかに組み入れるかといった点について検討・考察を行った。現在の提案手法については引き続き検討すべき点が大いに存在するが、AdaBoost をはじめとするアンサンブル学習と複雑適応系システムの統一的視座の構築の可能性を示す結果となった。

参考文献

- [Freund and Schapire, 1997] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119-139.
- [Booker et al. 1989] Booker, L. B., Goldberg, D. E., & Holland, J. H. (1989). Classifier systems and genetic algorithms. *Artificial intelligence*, 40(1), 235-282.
- [Farmer et al., 1986] Farmer, J. D., Packard, N. H., & Perelson, A. S. (1986). The immune system, adaptation, and machine learning. *Physica D: Nonlinear Phenomena*, 22(1), 187-204.
- [Zhou et al., 2014] Fang Zhou, Claire Q., & Ross. D. King (2014). Predicting the Geographical Origin of Music, *ICDM*
- [Lazarevic et al., 2002] Lazarevic, A., & Obradovic, Z. (2002). Boosting algorithms for parallel and distributed learning. *Distributed and Parallel Databases*, 11(2), 203-229.
- [Oza, 2005] Oza, N. C. (2005). Online bagging and boosting. In *Systems, man and cybernetics, 2005 IEEE international conference on (Vol. 3, pp. 2340-2345)*. IEEE.
- [Drucker, 1997] Drucker, H. (1997). Improving regressors using boosting techniques. In *ICML (Vol. 97, pp. 107-115)*.
- [Margineantu, 1997] Margineantu, D. D., & Dietterich, T. G. (1997). Pruning adaptive boosting. In *ICML (Vol. 97, pp. 211-218)*.