

機械学習を用いて賭け事を予測することは できるのか（仮）

清水 豪士

富山県立大学大学院 電子・情報工学専攻
u155016@st.pu-toyama.ac.jp

September 3, 2021

はじめに

背景

- 賭け事は娯楽としての印象が強い一方、保険はビジネスとして発展した賭け事の一種である。
- 賭け事を機械学習で予測できれば、安心して賭けることができるのではないか？
- 競馬には非常に豊富なデータがあり、機械学習の対象に適している。
- 特徴量となるものが多く、なにを使うかなどで予測結果が変化する。
馬名、騎手名、調教師、父馬、母父馬、競馬場、距離、出走履歴など

目的

- 機械学習を用いて、競馬予測をたてることができるモデルを構築する。

競馬

- 日本の競馬には JRA が主催する「中央競馬」と、地方公共団体が主催する「地方競馬」の 2 種類がある。
- 馬券の買い方には様々な種類があり、単勝、複勝、3 連複、3 連単などがある。

機械学習で競馬を採用するには

- 競馬を機械学習でやるにはどのような目的設定でやればいいのか.
- 1 荒れる競馬か荒れない競馬かを判別する.
 - 2 馬や騎手の強さを推定する.
 - 3 予測タイムを推定し、順位を予測する.
 - 4 出走馬の勝つ確率を求める.

アルゴリズム

SVM

- 機械学習の分野で最も広く使われているアルゴリズムの1つ.
- 特徴次元数が大きいデータでも高い精度がでるという性質をもつ.
- 非線形のデータも扱うことが出来る.

ニューラルネットワーク

- 人間の脳内にある神経細胞とそのつながりを人工ニューロンという数式的なモデルで表現したもの.
- 非線形学習ができる.

ランダムフォレスト

- 決定木を複数作成して、各決定木の出力結果の平均、多数決を算出して予測を行うモデル.
- 分類と回帰のどちらの問題にも利用できる.

アンサンブル学習

- 個々に別々の学習器として学習させたものを、融合させる事によって、未学習のデータに対しての予測能力を向上させるための学習のこと。
- 弱学習器を多く使うことで精度を上げる手法。

バギング

- 学習データのブーストストラップサンプルを利用して複数個の弱識別器の学習を行い、予測時にはそれら弱識別器の多数決で推論を行う手法。

ブースティング

- ブースティングとは複数の弱識別器を直列に学習し、学習された全てのモデルを用いて予測を行うアルゴリズムのこと。

ブーストラップ

- 訓練データからランダムに n 個のデータをサンプリングすること。

アルゴリズム

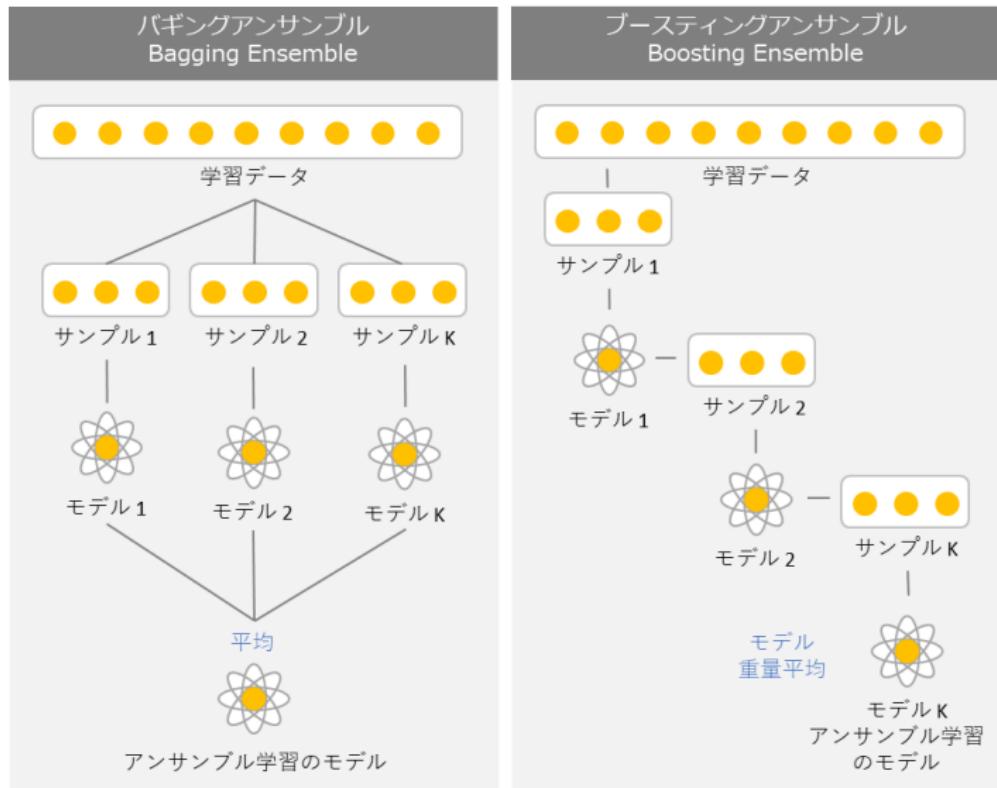


図 1: バギング

図 2: ブースティング

中間発表に向けて

リスト

- ① 既にやっている人の後をなぞる.
- ② スクレイピングなどの知識がないため、実践しながら勉強していく.
- ③ 論文などの参考文献、考え方の参考となるサイトなどを探す.
- ④ 分析手法について何が有効なのかを調べ、何を使用するかを考える.

スクレイピングの際に JRA の公式サイトからスクレイピングするのは困難なため、netkeiba というサイトからスクレイピングするようとする。
(必要な情報は netkeiba でも十分取得できる。)

まとめ

- 今回は修論の方向性、テーマを話した。

課題

- タイトル決め
- 実際に研究に着手する