

重み付け木を用いたランダムフォレストによる自動分類

小畑智広[†] 小林 学[‡] 渡辺重佳[‡]

[†]湘南工科大学大学院 [‡]湘南工科大学 工学部 情報工学科

1. はじめに

分類精度のあまり高くない弱学習器を複数用いて分類を行う集団学習法は、有効な手法であることが知られている。L.Breimanにより提案されたランダムフォレスト法[1](以下RF法と略す)は、ランダムに選択した少数の属性に対して未剪定の決定木を大量に生成し、各決定木による判別結果の多数決により自動分類を行う分類法であり、大変高い分類精度を達成する。一方モデルが未知のマルコフ情報源に対する優れた圧縮法として文脈木重み付け法(以下CTW法と略す)が提案されており[2,3],これを分類問題に直接適用した手法も提案されている[4]。CTW法は木構造における全ての確率モデルを混合する機構を持っているため、確率構造の選択を行う必要が無い。そこで本研究ではRF法により生成された各決定木に対してCTW法による確率モデルの混合によるクラス確率を求める手法(以下CTWRF法と略す)を示し、RF法との関連について考察を行う。

2. ランダムフォレスト(RF)法[1,5]

属性数を M とし、生成する木の総数を K とする。このときRF法のアルゴリズムを次に示す。

- ① 与えられたデータからランダムにサンプルを選択し K 組のブートストラップサンプル集合 B_1, B_2, \dots, B_K を作成する。
- ② サンプル集合 $B_k (k = 1, 2, \dots, K)$ それぞれに対し、CARTアルゴリズムを用いて B_k に対する決定木 T_k を生成する。ただしCARTにおけるノードの展開時に候補となる属性は、 M 個存在する全属性の中から m 個の属性をランダムに抽出し、この m 個の属性に対してのみ選択基準を用いて属性選択を行う。ただし $m \ll M$ とする。分類問題の場合 $m = \sqrt{M}$ が用いられることが多い。
- ③ 生成した全ての T_k に対して新規テストデータに対する判別結果を求め、最も判別結果の多いカテゴリに分類する。□

RF法は、ランダムに選択された少数の属性による弱学習器を多数用いることにより、過学習を防ぎつつ精度良い分類器の生成に成功している。

3. 重み付け木を用いたRF(CTWRF)法

決定木に対してCTW法を適用するために、RF法のステップ②を以下のように修正する。

- ② サンプル集合 $B_k (k = 1, 2, \dots, K)$ それぞれに対し、 M 個の全属性の中から m 個の属性をランダムに抽出する。ただし $m \ll M$ とする。 B_k に対してこの m 個の属性のみを選択対象としたCARTアルゴリズムを用いて決定木 T_k を生成する。

この修正ステップ②により、 T_k 中には合計 m 個以下の属性しか出てこない点に注意されたい。

次に各決定木 T_k に対し、CTW法を応用した分類アルゴリズムを示す。本節では簡単のためカテゴリが2値(0 or 1)の場合について述べる。ただし、多値の場合への拡張は容易である。あるサンプル集合 B_k について考える。前節で述べたRF法のステップ②で生成された T_k のあるノード s について、そのノード s の下での B_k 中のカテゴリ0のデータ数を a_s 、カテゴリ1のデータ数を b_s とする。このときKT推定量を用いてノード s の下での B_k の確率を

$$P_e(B_k|s, T_k) = P_e(a_s, b_s) = \frac{1/2 \cdots (a_s - 1/2) \cdot 1/2 \cdots (b_s - 1/2)}{1 \cdot 2 \cdots (a_s + b_s)} \quad (1)$$

とする。次に T_k の下での B_k の確率を推定することを考える。まずノード s が T_k の葉のときには $P_w^s = P_e(a_s, b_s)$ とする。ノード s に対して分岐した子ノード $0s$ 及び $1s$ に対する重み付け確率を P_w^{0s} 及び P_w^{1s} としたとき、 s における重み付け確率を以下のように再帰的に定義する[2,3,4]。

$$P_w^s = \alpha P_e(a_s, b_s) + (1 - \alpha) P_w^{0s} \cdot P_w^{1s} \quad (2)$$

結果的に根ノード λ に対する P_w^λ を $P_w(B_k|T_k)$ と考える。

次に新規テストデータ \mathbf{x} に対応する未知のカテゴリに対する確率変数を y とする。このとき $D_k^{(y)} = B_k \cup \{(\mathbf{x}, y)\}$ に対して上と同様の方法で $P_w(D_k^{(y)}|T_k)$ を求める。結果的に

$$P_w((\mathbf{x}, y)|T_k, B_k) = P_w(D_k^{(y)}|T_k) / P_w(B_k|T_k) \quad (3)$$

と推定し、 T_k, B_k における (\mathbf{x}, y) の理想符号長を

$$L(\mathbf{x}, y|T_k, B_k) = -\log_2 P_w((\mathbf{x}, y)|T_k, B_k) \quad (4)$$

と計算する。本研究では簡単のため T_k 及び B_k に対する確率は全て同じと仮定し、

$$L(\mathbf{x}, y) = \frac{1}{K} \sum_{k=1}^K L(\mathbf{x}, y|T_k, B_k) \quad (5)$$

とする。結果的に \mathbf{x} の推定カテゴリは以下とする。

$$\hat{y} = \operatorname{argmin}_y L(\mathbf{x}, y) \quad (6)$$

A Classification Method Based on Random Forests Using Context Tree Weighting Method
Tomohiro OBATA[†], Manabu KOBAYASHI[†] and Shigeyoshi WATANABE[†]

[†]Graduate School of Engineering, Shonan Institute of Technology, [‡]Shonan Institute of Technology

式(6)が示すように、 $P_w(D_k^{(y)}|T_k)$ はカテゴリ数だけ計算する点に注意が必要である。ここで $P_w(B_k|T_k)$ は予め学習時に計算しておくことができる。また

$P_w(D_k^{(y)}|T_k)$ は、 $P_w(B_k|T_k)$ の計算過程を利用して容易に求めることが可能である。このように、CTW法の効率的なアルゴリズムを利用してカテゴリの分類を行うことが可能である。さらに提案手法は、文献[2,3]のCTW法に対するフリーズを導入し、より効率的に計算することも可能である。

4. 計算機実験による評価及び考察

本節では毎日新聞の記事データ[6]を用いて、2節及び3節の分類手法に対する評価を行う。実験では新聞のカテゴリとして「文化」、「芸能」、「経済」、「国際」、「スポーツ」の5カテゴリを対象とし、各カテゴリに対して学習文書と新規テスト文書をそれぞれ用意した。なお学習文書及び新規テスト文書は重複が無いようにランダムに選択し、10回試行を行った平均の誤分類率で評価する。評価のための新規テスト文書数は、各カテゴリに対して400ずつとする。

実験では各文書に対してChaSen[7]を用いて形態素解析を行い、助詞、助動詞を除いた単語を抽出する。次にそれぞれの文書に対して各単語が出現していれば1、していなければ0とした文書ベクトルを構成する。抽出された総単語数が属性数に対応しており、 $M = 68328$ である。これらの文書ベクトルを用いてRF法及びCTWRF法に対して評価を行う。

CTWRF法について予備実験を行った結果 $m = M^{0.75}$ とした場合の性能が最も良かった。そこで以下CTWRF法では $m = M^{0.75}$ を用いた。

決定木の数及び学習文書数によって性能がどの程度異なるかを評価する。ここで各カテゴリに対して、学習文書数をすべて同一の($N_i = 10, 50, 100$)とし、決定木を($K = 10, 50, 100$)とした。それぞれの結果を表1, 2, 3に示す。通常のRF法では $m = \sqrt{M}$ を用いることが多いが、本実験では $m = M^{0.75}$ の場合も合わせて示した。またCTWRFに対し、 $\alpha = 0.5, 0.9, 1.0$ の結果を示している。

表の結果を見ると、CTWRF法は $\alpha = 0.9, 1.0$ のときにRF法に比べ誤分類率が低くなった。一方 α が小さくなると誤分類率が増えてしまっている。CTW法では α が小さいほど小さな木の確率モデルのウェイトを上げる役割を持つ。従って大きな確率モデルを用いることがCTWRF法において良い結果を示すことを意味する。これはRF法においてCARTの出力の木に対して枝刈りを行わずに大きな木をそのまま利用することに該当する。

CTW法は小さなモデルほどノードに割り振られたパラメータの推定精度が上がる良さを持っているが、今回の実験ではそのような傾向を確かめることはできなかった。パラメータの推定精度との関連については今後より詳細な検討を行い、明らかにしたい。

表 1: $N_i = 10$ の場合の誤分類率(%)

K	10	50	100
RF法($m = \sqrt{M}$)	61.9	51.0	48.2
RF法($m = M^{0.75}$)	57.0	49.1	46.7
CTWRF($\alpha = 0.5$)	71.7	70.4	67.9
CTWRF($\alpha = 0.9$)	56.2	47.2	46.7
CTWRF($\alpha = 1.0$)	56.2	47.3	46.7

表 2: $N_i = 50$ の場合の誤分類率(%)

K	10	50	100
RF法($m = \sqrt{M}$)	38.9	26.1	23.8
RF法($m = M^{0.75}$)	32.9	26.0	25.2
CTWRF($\alpha = 0.5$)	47.1	43.2	43.1
CTWRF($\alpha = 0.9$)	31.5	22.4	20.8
CTWRF($\alpha = 1.0$)	31.5	22.5	20.8

表 3: $N_i = 100$ の場合の誤分類率(%)

K	10	50	100
RF法($m = \sqrt{M}$)	32.1	21.7	20.3
RF法($m = M^{0.75}$)	25.8	20.7	20.0
CTWRF($\alpha = 0.5$)	41.6	38.1	36.7
CTWRF($\alpha = 0.9$)	25.1	16.6	15.3
CTWRF($\alpha = 1.0$)	25.1	16.7	15.4

5. まとめと今後の課題

本研究ではRF法により生成された各決定木に対してCTW法による確率モデルの混合によるクラス確率を求める手法を示した。新聞記事データを用いた計算機実験では、RF法に比べ高い正分類率を示すことが分かった。今後パラメータの推定精度と決定木の大きさの関係について明らかにしたい。

謝辞

本研究の一部はJSPS 科研費基盤研究(C) No.25330045の助成を受けたものです。

参考文献

- [1] L. Breiman. Random forests. *Mach. Learning*, 45(1):5-32, 2001.
- [2] F.M.J. Willems, Y.M. Shtarkov and T.J. Tjalkens, "The Context Tree Weighting Method: Basic Properties," *IEEE Trans. On Information Theory*, vol.41, no.3, pp.653-664, May 1995.
- [3] Z. Dawy, J. Hagenauer and A. Hoffmann, "Implementing the context tree weighting method for context recognition," *Proc. Of the IEEE Data Compression Conference*, p.536, March 2004.
- [4] 小畑智広, 池上裕之, 小林 学, 坂下善彦, "文脈木重み付け法による確率モデルを限定した文書分類", *電子情報通信学会論文誌(D)*, Vol.J95-D, No.10, Oct. 2012.
- [5] 後藤正幸, 小林学. 入門 パターン認識と機械学習. コロナ社, 2014.
- [6] CD-毎日新聞94'データ集, 日外アソシエーツ, 1995.
- [7] 奈良先端科学技術大学院大学 ChaSen-形態素解析器 <http://chasen-legacy.sourceforge.jp/>