ISS-A-040

勾配ブースティング型アンサンブル学習を用いた母親の育児対応判定

Determining mother's response to childcare using ensemble learning with gradient boosting

† 阪野 ヒカル　Hikaru BANNO
† 川口 友稀　Yuuki KAWAGUCHI
† 藤部 虎大　Kota FUJIBE
† 田邉 造　Nari TANABE

公立諏訪東京理科大学 †Suwa University of Science
E-mail:†{t119113@ed, t118158@ed, gh21520@ed, nari@rs}.sus.ac.jp

1　はじめに

本論文は，母親の育児対応をアンサンブル学習によって判定する手法を提案する．提案手法は，(Step 1) バターワースフィルタと有色駆動源カルマンフィルタにより乳児が泣いた時の腸音を抽出する．次いで (Step 2) 抽出された腸音と体動の生態特徴量から勾配ブースティング型アンサンブル学習を用いて，母親の育児対応を判定している．提案手法の特徴は，母親の育児の一助となることである．

2　提案手法

Step 1 腸音抽出 [1]

観測信号に腸音抽出で不要となる定常雑音スペクトルが存在するため，バターワース型の LPF と HPF を用いて定常雑音スペクトルを除去し，高周波と低周波数領域のノイズをカットする．

次に，時刻 $n+1$ における観測信号から腸音信号を推定するために，以下の状態空間モデルを解く必要がある．

$$\begin{cases} \boldsymbol{d}(n+1) = \Phi \boldsymbol{d}(n) + \boldsymbol{\delta}(n+1) \\ \boldsymbol{s}(n+1) = M^T \boldsymbol{d}(n+1) + \boldsymbol{\epsilon}(n+1) \end{cases} \tag{1}$$

ここで，$\boldsymbol{d}(n+1)$ は腸音信号ベクトル，Φ は状態遷移行列，$\boldsymbol{\delta}(n+1)$ は有色駆動源ベクトル，$\boldsymbol{s}(n+1)$ は観測信号ベクトル，M は観測遷移行列，$\boldsymbol{\epsilon}(n+1)$ は雑音信号ベクトルである．

以上の状態空間モデルを有色駆動源カルマンフィルタで解くことで，腸音信号を推定する．

Step 2 生体特徴量

腸音量や体動量の生態特徴量は，乳児の泣き始めから泣き終わりまでの 1 分ごとの平均量を基本データセットとする．また，このときに母親が取った対応で乳児の機嫌が改善されたかどうかを正解ラベルとする．

Step 3-1 決定木分析

決定木のデータ分類は，データの散らばりの指標であるエントロピー E

$$E = -p_t \log_2 p_t - p_f \log_2 p_f \tag{2}$$

を用いて分類される．ここで，p_t と p_f は正しく識別されたデータの割合と間違って識別されたデータの割合を表す．

分類後のデータのエントロピー E_{after} が分類前のエントロピー E_{before} よりも小さくなれば，分類によってデータの散らばりが纏まりデータが整理される．このことから，情報獲得量 $Gain$ は

$$Gain = E_{before} - E_{after} \tag{3}$$

と定義される．正しく分類された $Gain$ の多いノードを順に選択することで，高い識別能力を持った木構造の識別モデルが取得可能となる．

Step 3-2 勾配ブースティングによるモデル追加 [2]

勾配ブースティングによるデータの識別において，識別の正誤を評価するための損失関数 $L(y, F(x))$ は

$$L(y,F(x)) = -y \times (\log(\frac{1}{1+\exp F(x)}) - log(1 - \frac{1}{1+\exp F(x)})) - log(\frac{1}{1+\exp F(x)}) \tag{4}$$

と定義される．ここで，y はデータラベル，x は学習用データ，$F(x)$ は決定木の出力である．データラベル y は乳児の機嫌が良い時に 1，機嫌が悪い時に 0 を取る．

損失関数 $L(y, F(x))$ が最小となるような勾配 γ を持つ弱識別器 $h(x)$ は

$$h(x) = \arg\min_{\gamma} \nabla_F L(y, F(x)) \tag{5}$$

と表せる．このことより，損失関数 $L(y, F(x))$ が最適化された決定木の出力 $F(x)$ は学習率 α を用いて

$$F(x) = \alpha h(x) \tag{6}$$

となる．

生成された決定木の出力 $F(x)$ において，式 (5) により最適な損失関数 $L(y, F(x))$ を再び考え，新たな弱識別器 $h(x)$ を求める．最終的な決定木の出力 $F_M(x)$ は，生成された $h(x)$ を逐次加えることで

$$F_M(x) = \alpha \sum_{m=1}^{M} h_m(x) \tag{7}$$

となる．

3　計算機シミュレーション

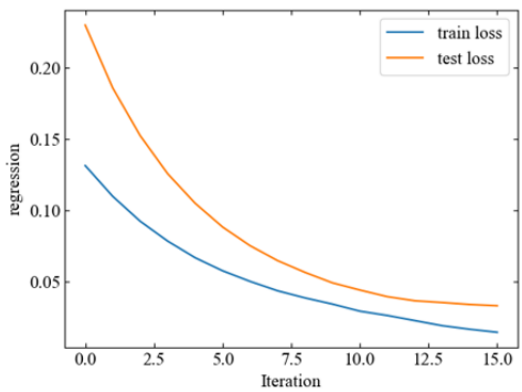

図1：誤差関数Lの推移

図 1 は乳児の生体特徴量と育児対応のデータを用いた勾配ブースティングの結果である．提案手法の識別精度は約 85% であり，乳児の生体特徴量と母親の育児対応には関連性があると考えられる．また，乳児の生体特徴量から母親の育児対応も予測して判定することが可能であると考えられる．

4　まとめ

本論文は，生体特徴量を用いて母親の育児対応の正誤を予測することを提案した．勾配ブースティングによる識別精度は 85% となることから，乳児への対応の正誤判定は可能といえる．

参考文献

[1] 荒木雅弘 ' フリーソフトではじめる機械学習入門' 森北出版株式会社 2017 年 11 月

[2] Nari TANABE, Toshihiro FURUKAWA, Shigeo TSUJII"Robust Noise Suppression Algorithm with the Kalman Filter Theory for White and Colored Disturbance," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Volume E91.A, Issue 3, pp. 818-829 (2010).