アンサンブル学習法の河川景観満足度調査データ解析への応用

下川　敏雄[1]・武藤由香里[2]・御園生拓[3]・北村　眞一[4]

[1]非会員　博士(工学)　山梨大学大学院医学工学総合研究部（〒400-8511　山梨県甲府市武田4-3-11 ,E-mail; shimokawa@yamanashi.ac.jp）
[2]学生会員　山梨大学大学院医学工学総合教育部持続社会形成専攻（〒400-8511　山梨県甲府市武田4-3-11 ,E-mail;g07mf014@yamanashi.ac.jp）
[3]非会員　Ph.D in Biology　山梨大学大学院医学工学総合研究部（〒400-8511　山梨県甲府市武田4-3-11 , E-mail; mist@yamanashi.ac.jp）
[4]正会員　工博　山梨大学大学院医学工学総合研究部（〒400-8511　山梨県甲府市武田4-3-11 E-mail;skita@yamanashi.ac.jp）

近年，河川開発と景観に関する多くの報告が行われている．その調査の多くがアンケートに基づいており，その結果は重回帰分析のような線形モデルによって処理される．ただし，真の構造が単純な線形関係で得られることは殆どなく，非線形構造，および交互作用を伴うことが多い．本論文では，これらの複雑な構造を適切に捉えるだけでなく，その結果をグラフィカルに解釈できるアンサンブル型学習法の適用について述べた．そして，その有用性は河川景観満足度に関するアンケート調査[17)]への適用により提示した．その結果，アンサンブル型学習法を適用することで，重回帰分析，あるいは樹木構造接近法に比べて良好なモデル適合度をもつだけでなく，より有用な知見を見出すことができた．

Key Words : *Multivariate Additive Regression Trees, Interaction Effect, Variable Importance, Nonlinear Regression Models*

1．序

近年，河川開発と景観に関する多くの報告が行われている．またそうした研究を元に河川景観の設計に関する著作やマニュアル類が出版されている．

例えば，近年の和田他[18)]による大正川に関する調査では，河川の改善事業として望むことは，水質の改善，豊かな自然，親水空間の演出であることが報告されている．また，小路他[5)]は，緑，人工構造物，景観障害物が，景観評価に与える影響が大きいことを提示している．さらに，市民満足学会・（株）ワード研究所[17)]は，インターネット上での河川景観満足度に対する大規模調査を行い，重相関分析により河川景観満足度に与える影響を調査した．その結果，重要な要素として自然，うるおい，町並み・田園との調和，統一感，水質，堰や橋，安全が挙げられたことを報告している．河川景観の研究を基礎にした著作には，例えば土木学会編[7)]，島谷幸広編著[4)]などがあり，マニュアル類としては，建設省中国地方建設局太田川工事事務所[3)]，河川景観の形成と保全の考え方検討委員会編[2)]などが代表的であるが，これらはいずれも河川景観のデザイン面からの技術を提供するものである．

アンケート調査だけでなく，多くの回帰問題においては，一般に重回帰分析および，その他の線形モデル，例えば主成分回帰，部分最小2乗法などが頻用されている．線形モデルは説明変数あるいはその合成変数の線形結合によりモデルが構成されるため，結果の解釈は平易である．ただし，実際の事象はこのような単純なモデルで捉えられることは少なく，非線形構造をもつことが多い．さらに，説明変数間の交互作用関係(相乗関係)は予めモデルに組み込まなければならない．

一般に，回帰分析の目標は，(1)応答の良好な予測ができること，および(2)応答と説明変数の関連性(推定モデルの解釈)を明らかにすること，の二つに分けて捉えられるが，線形モデルが目標(1)を充足しているとは言い難い．当てはまりの粗悪なモデルによる解釈は，誤った結論を導く惧れがある．

目標(1)に重点をおくのであれば，一般化加法モデル[8)]，ニューラル・ネットワーク[16)]などの適用が考えられる．ただし，これらの方法では，影響要因と結果のあいだの関係がブラックボックス化されるため目標(2)が不十分になる．

これに対して，目標(2)に重点をおいた重回帰分析の代替手法として，樹木構造接近法(あるいは決定木)が注目されている．この方法の利点は，モデルの非線形性あるいは交互作用関係を，解釈が平易なプロダクション・ルールによって与えることができることにあり，その有用性は多くのデータ・マイニングの成書で指摘されている(例えば，Witten & Frank[1] を参照).

ただし，樹木構造接近法による近似はステップ関数に基づくため，真のモデルが線形構造を持つ場合には不適切な結果を導く．また，一般に樹木構造接近法の予測確度は低いことも指摘されている[9].

これらの二つの目標を達成することは，アンケート調査の結果に内在する複雑な構造を適切に捉えるだけでなく，そこに新たな知見を見出すことに強く寄与すると考えられる．そのための一つの戦略は，目標(2)を満たす樹木構造接近法のモデル適合度を向上させ，目標(1)を満たすことである．本論文では，このような統計的学習法として，アンサンブル学習法をとり上げる．アンサンブル学習法とは，複数個の樹木モデルの加法型でモデルを構築することで，樹木構造接近法の利点を保持しながらモデル適合度を向上することができる方法である．また，変数重要度や部分従属度といった統計量を用いることで，影響要因と結果の間の関係を明らかにすることもできる．これにより，これまでのアンケート調査の分析では得られなかった，非線形関係，あるいは要因の影響の大きさが高い予測性能のもとで解釈できる．

ここでは，ワード研究所の河川快適性に関するアンケート調査をもとに，影響要因の評価を行う．アンサンブル型学習法のなかでも，とくに多重加法型回帰樹木(MART法)[10]に注目し，重回帰分析，樹木構造接近法，あるいはその他の回帰手法との性能を比較することにより，MART法によるデータ分析がこれまでの方法よりも，より多くの知見を与えることを示す.

２．アンケート調査に対する諸種の統計的方法とその問題点

一般に，本研究で扱う景観評価の要因分析は，提示する対象のイメージ，操作できる要因の設定，アンケート対象者あるいは実験の被験者の 相違によって，種々異なる結果を得ている．操作要因は大別して，物理的要因，画像的要因，心理的要因に分けられる．物理的要因としては，河川の自然の保全されている状況（水質，植生など）や建設されている人工構造物（護岸の材料や形態など）が挙げられる．画像的要因は，評価する写真などの特性で，色彩の組み合わせと面積配分，電線や物体の輪郭線などの錯綜度や統一度やフラクタル次元などである．心理的要因は，景観から受ける心理的尺度の反応で，整然性，活動性，親近性，賑わい感などである．これらをまとめると，河川景観の評価は，（１）河川の物理的特性，（２）河川を見る視点すなわち写真の取り方による画像的特性，（３）河川から受ける感性的印象の複合的影響に依るものと考えられる.

河川イメージの提示では，河川の現場の写真，CGによる合成画像，河川の特定の場所の記憶（地名や河川名）などが用いられている．アンケート対象者は，限られた地域の住民が多く，その数は数百人から数千人，被験者は学生や研究所などの職員，人数も数十人程度にに限定され，それぞれの操作要因の限定，現状の河川の範囲，CG表現の限界，被験者特性などによる偏りが生じている．今回用いた（株）ワード研究所の調査データは，実際の現場の河川を対象とし，河川の景観の記憶に基づく判断，回答者数は十分多いが，登録されたアンケート対象者の限定といった限界がある.

こうしたデータの制約に基づく適用可能性と普遍性の限界があるが，これらは研究事例を重ねていくことによって普遍的真理へアプローチするものと考えられる.

計量データから，順序データ，カテゴリーデータ等によって要因を分析する統計的手法には，線形の回帰分析として代表的な重回帰分析から，順序データを扱うコンジョイント分析，カテゴリカルな回帰分析としての林の数量化理論1類・2類などがあるが，これらの分析手法の問題点としては，(1)尺度化の問題（人間は選好順位，間隔尺度，比率の判断が可能であるか，またその信頼性の程度はどのくらいか，(2)多変量解析などの統計的分析手法の問題は，非線形性，変数の適用範囲，予測の精度とその評価（適合度，的中率など）をどうするか，などが挙げられる.

３．樹木構造接近法

樹木構造接近法は，モデルへの適用結果が「樹木」によってグラフィカルに表現される．樹木表現は，複雑な非線形効果や交互作用効果に対する鋭い洞察を与えるだけでなく，さらに，複数の尺度が混在した場合にも平易な解釈ができる．分類回帰樹木(CART:Classification And Regression Tree)[14]は，樹木構造接近法を飛躍的に進歩させ，社会科学・工学・医学などの諸種の応用分野に広く関心を与えている[6].

(1)CART法

CART法は，データを説明変数空間に沿って2分岐させることで，モデルをあてはめる．このとき，分岐させた部分集合(ふし)内の応答の予測値には，平均値あるいは中央値が用いられる．したがって，CART法では，ステップ関数が当てはめられる．そのモデル構築の過程は，(1)分岐過程，(2)刈り込み過程，(3)最適樹木の選定過程，

から成る．

分岐の評価基準には，2分岐されたそれぞれの部分集合の応答に対する不均一性の測度が用いられる．いま，応答 $\{y_i\}_{i=1}^n$ とそれに対応する p 次元説明変数ベクトル $\{\boldsymbol{x}_i\}_{i=1}^n$ が与えれたとき，任意の部分集合(ふし)t に属する $|t|$ 個の個体の応答 y に関する平方和は

$$\varphi(t) = \sum_{i \in t} \{y_i - \bar{y}(t)\}^2$$

である．ここに，$\bar{y}(t)$ は，t に含まれる個体の応答の平均値(あるいは中央値)である．分岐過程では，上式が最小となる最適分岐変数および分岐点を探索し，逐次に分岐を続ける．この分岐の仮定は樹木形式のグラフィクスで表示することが多く，これが樹木構造接近法の名前の由来になっている．この樹木の成長過程は，任意の停止基準に到達するまで続けられる．分岐過程で得られる樹木は，一般に過剰適合を示していることが多く，以後の過程ではモデル適合度と簡便さのトレードオフを意図して実行される．

刈り込み過程では，分岐過程により得られた過剰適合な樹木を，刈り込み基準(複雑度コスト)に基づいて，根幹ふし(分岐のない状態)まで，分岐点を逐次に削除する．これにより，大きな樹木から根幹ふしまでの巣籠もり状の樹木系列が得られる．

そして，部分樹木系列のなかから残差平方和推定値(例えば交差確認推定値)を用いて，これが最小になる部分樹木を最適モデルとして選択する．

CART法により得られたモデルは

$$h(\boldsymbol{x}; \{t_m\}_{m=1}^M) = \sum_{m=1}^M \hat{\beta}_m \mathrm{I}(\boldsymbol{x}_n \in t_m) \qquad (3.1)$$

で与えられる．ここに，$\hat{\beta}_m = \bar{y}(t_m)$ であり，I(·) は括弧内が真なら1，偽なら0を返す指標関数であり，$m(=1,\ldots,M)$ は終結ふし(最終の部分集合)を表す．また，パラメータの推定はふし内の観測値の平均値あるいは中央値で行われる．

CART法の利点を以下に示す：(a) 分岐点の探索は，全ての説明変数の分岐候補を評価するだけなので，外れ値の影響が小さいだけでなく，多重共線性の影響を受けない，(2)説明変数の数が標本サイズを上回る場合でもCART法は適用できる，(3) 説明変数に対する単調変換に対して不変であるため，重回帰分析で用いられる変数変換(例えば対数変換)を必要としない．(4) 変数の尺度あるいは欠測値を考慮しなくても適用できる．したがって，アンケート調査のような複数の尺度があるデータ解析に適している．すなわち，前節で挙げた既存の統計分析手法の問題点をCART法は回避することができる．

(2) 変数重要度

CART法のもう一つの利点は，ある応答に対して各説明変数がどの程度の重要度をもつかを統計量により明らかにできる点にある．ただし，得られたCART樹木のみに基づく解釈だけでは，分岐に用いられていない説明変数の効果は隠される．例えば，ある説明変数は最初の分岐で2番目に良い分岐ルールを持ったが，最終的には樹木分岐として採用されなかったとする．この場合に，その変数の重要度を0とすることが必ずしも適切であるとは限らない．そのため，変数重要度の算出方法には若干の工夫が行われている．すなわち，CART樹木の変数重要度は，得られたCART樹木の分岐変数とは異なる変数(代理変数)で分岐を行ったときの残差の減少量(改善度)を用い，この改善度を(分岐過程で得られた)樹木の全ての分岐点で計算する．そして，それぞれの変数の重要度は，これらの改善度の総和により定義される．

通常，変数重要度は，最大の変数重要度をもつ変数の値を100としたときの相対指標として解釈される．

4．アンサンブル型学習法

線形回帰法は，説明変数と応答の非線形構造を適切に捉えることができないだけでなく，説明変数間の交互作用効果を事前にモデルにとり込まなければならない．他方，CART法では非線形構造および交互作用効果を解釈が平易な樹木によって提示できるものの，その近似はステップ関数に基づくため，真のモデルが線形構造を持つ場合には不適切な結果を導く[14]．さらに，CART法のモデル適合度は低いことも二三の論文で指摘されている[6]．近年，諸種のアンサンブル学習法，例えばBagging法[11]やRandomForest法[13]が提案されている．

アンサンブル学習法とは，線形回帰モデルあるいは，CARTモデルといった学習器を反復してあてはめることで，より強力な予測力をもつモデルを生成する統計的学習法の一つである．

とくに，CART樹木をアンサンブル結合するブースティング 樹木法は，CARTの長所を保持しながら，その欠点の予測性能を向上することに成功している．これは，MART法(Multiple Additive Regression Trees: 加法型重回帰樹木)と名づけられている[6]．

(1) MART法

MART法は，CART法の予測確度を向上させる目的で，

CART樹木にブースティングを加味する方法として考案された．ブースティングとは，関数空間における最適化アルゴリズムとして提案された方法である[12]．また，ブースティング法の非常に興味深い側面は，データの背後にある潜在モデルに含まれる主効果項あるいは交互作用項のような構造的な推定問題にも魅力的な効用を発揮することである．このことは，アンケート調査のような複雑な構造を適切に捉えることに寄与すると考えられる．

観測値$\{y, \boldsymbol{x}\}$が与えられたとき，MART法によるブースティング法の目標は，(微分可能な)損失関数

$L(y, f(\boldsymbol{x}))$の期待値を最小にするモデル

$$f^*(\boldsymbol{x}) = \arg\min_{f(\boldsymbol{x})} E[L(y, f(\boldsymbol{x}))] \qquad (4.1)$$

を推定することである．本論文では，線形回帰法あるいはCART法と同様に，損失関数には，最小2乗基準損失関数

$$L(y, f(\boldsymbol{x})) = (y - f(\boldsymbol{x}))^2 \qquad (4.2)$$

を用いる．

B個のCART樹木$h(\boldsymbol{x}; \{t_m^{(b)}\}_{m=1}^{M_b})$より得られる，MART法の推定モデル$\hat{f}_{\text{boost}}^{(\text{B})}(\boldsymbol{x})$は，加法的な展開のもとで逐次的に式(4.1)を近似することで，

$$\begin{aligned}\hat{f}_{\text{boost}}^{(\text{B})}(\boldsymbol{x}) &= \sum_{b=1}^{B} v\hat{\omega}_b h(\boldsymbol{x}; \{t_m^{(b)}\}_{m=1}^{M_b}) \\ &= \sum_{b=1}^{B} v \sum_{m=1}^{M_b} \hat{\omega}_b \hat{\beta}_m^{(b)} \mathrm{I}(\boldsymbol{x}_n \in t_m^{(b)})\end{aligned}$$

により得られる．ここに，$\hat{\omega}_b$は，b番目のCART樹木での重みパラメータの推定値であり，$v(0 < v \leq 1)$は，学習効率を制御するための任意の縮小パラメータである．

このとき，重みパラメータおよび樹木の推定を同時に行うことは困難なため，MART法でのブースティングでは，重みパラメータω_bを推定する過程と，樹木$h(\boldsymbol{x}; \{t_m^{(b)}\}_{m=1}^{M_b})$を推定する過程を交互に繰り返す反復アルゴリズム(ステイジワイズ戦略)を採用している．

通常，回帰モデルの推定において，「ステップワイズ」戦略が頻用されている．ステップワイズ戦略では，変数(あるいは基底関数)が追加される毎に，モデルに含まれる全てのパラメータが調整されるのに対して，ステイジワイズ戦略では，パラメータの調整は必要としない．さらに，ステイジワイズ過程では，これらの推定を最急降下法に類似する更新手続きとして捉え，損失関数(4.2)に対してCART樹木を当てはめるのではなく，その偏微分したもの(これを疑似残差と呼ぶ)に対して当てはめる．

(2) 変数重要度の拡張

MARTモデルが，複数のCART樹木の加法形であることから，変数重要度は，CART樹木の場合と同様に定義できる．ただし，CART樹木における代理変数による変数重要度の定義は採用しない．もともと，CART法で代理変数を用いた理由は，単一の樹木の分岐でマスキングされた変数の影響をできる限り公平に要約することにある(3.2節参照)．これに対して，MART法では多数の樹木を扱うが，このことは単一の樹木におけるマスキング変数の影響を考慮することにも繋がる．したがって，MART法での変数重要度は，推定された複数のCART樹木での改善度(それぞれのCART樹木の構成に出現しなかった変数の改善度は0である)の算術平均値によって定義される．

(3) 部分従属度

MART法では，応答と説明変数のあいだの関数関係を明示できない．そのため，Friedman[10]は，これらの関係をグラフィカルに捉えるための方法，すなわち，部分従属プロットを提案している．

いま，関心がある$p^+(p^+ = 1,2)$個の説明変数を$\boldsymbol{z}_i$とし，それ以外を$\boldsymbol{z}_i^c$とする．すなわち，$\boldsymbol{x}_n = z_n \cup z_n^c$である．したがって，推定されたMARTモデル$\hat{f}_{\text{boost}}^{(\text{B})}(\boldsymbol{x})$を$\hat{f}_{\text{boost}}^{(\text{B})}(z, z^c)$と書くことができる．このとき，ある固定された$\boldsymbol{z}^c$のもとで$\hat{f}_{\text{boost}}^{(\text{B})}(\boldsymbol{x} \mid \boldsymbol{z}^c)$が$\boldsymbol{z}^c$のバラツキに強く影響を受けないという仮定すると，$\boldsymbol{z}_i$の部分従属度[10]は

$$D(z_i) = \int \hat{f}_{\text{boost}}^{(\text{B})}(z_i, z^c) d\Pr(z^c) \qquad (4.3)$$

で定義できる．ここに，$\Pr(z^c)$ は z^c の周辺確率密度である．実際には，式(3.3)の経験推定値

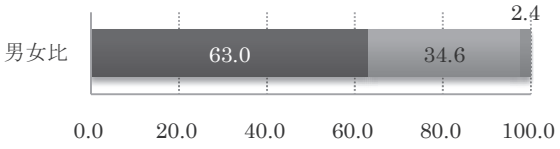
$$\hat{D}(z_i)=\sum_{l=1}^{n}\hat{f}^{(\mathrm{B})}_{\mathrm{boost}}(z_i,z_l^c)/n$$

を用いる．このとき，部分従属プロットは，座標軸上に $\left(z_i,\hat{D}(z_i)\right),i=1,2,\ldots,n$ をプロットし，それぞれを折れ線グラフによって結びつけることで構成される[6]．

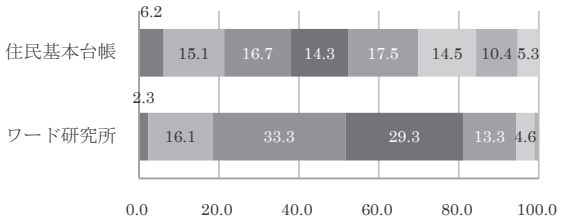
6．河川景観満足度調査に対するアンサンブル型学習の適用

河川景観の満足度の現況を調査する目的で，平成17年6月30日から7月24日にかけて，㈱ワード研究所によってアンケートが実施されている．このアンケートは，全国を対象にしたインターネット調査の方法で行われており，12,189名の回答が得られた．本データの解析の目標は，河川の満足度に対する影響は何かを探索することにある．

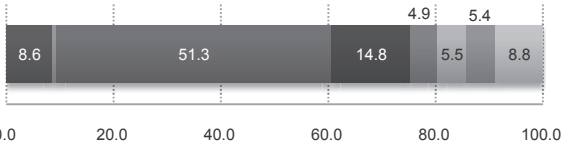
回答者の背景を図-1に提示する．男女比は，男性のほうが女性に比して約30%程度高かった (n=12189)．年代では，30代から40代にかけての比率が住民基本台帳に比して高く，他方，60代以上の高齢者の比率が低かった(n=11509)．さらに職業別では，会社員・団体職員・公務員が半数を占めており，また，主婦の比率



(a) 性別(棒グラフは，左から順に，男性，女性，不明で構成されている)



(b) 年齢(棒グラフは，左から順に，10代，20代，30代，40代，50代，60代，70代(1.0%)，80代以上(0.4%)で構成されている)



(c)職業(棒グラフは，左から順に，自営業主，家族従業者(0.7%)，会社・団体職員・公務員(パート含む)，主婦，学生，無職，その他，不明で構成されている)

図-1 河川景観満足度アンケートの要約

表-1：アンケート項目に対する相関行列

	水量	自然度	水質	ゴミ	四季	親水性	安全性	生態系	橋・堰	コンクリート	応答
水量	1.000										
自然度	0.409	1.000									
水質	0.340	0.664	1.000								
ゴミ	0.237	0.557	0.706	1.000							
四季	0.302	0.513	0.504	0.479	1.000						
親水性	0.280	0.441	0.461	0.421	0.506	1.000					
安全性	0.176	0.389	0.477	0.496	0.442	0.559	1.000				
生態系	0.196	0.510	0.518	0.439	0.478	0.439	0.433	1.000			
橋・堰	0.284	0.413	0.426	0.425	0.473	0.432	0.454	0.427	1.000		
コンクリート	0.217	0.565	0.464	0.463	0.399	0.366	0.350	0.447	0.445	1.000	
応答	0.342	0.614	0.603	0.578	0.594	0.548	0.532	0.509	0.550	0.519	1.000

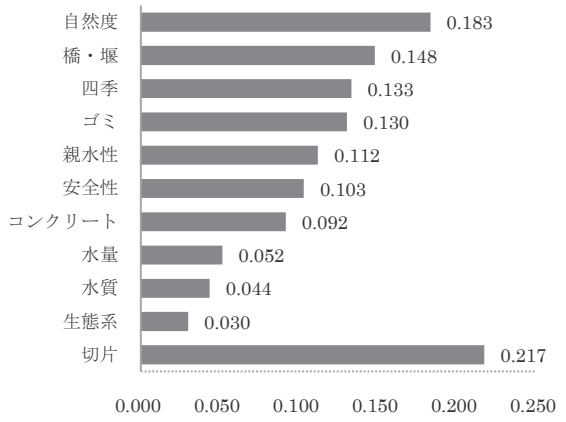


図-2 重回帰分析の回帰係数

も14.8%であり，2番目に高かった (n=12189)．これは，本アンケートがインターネットで実施されていることから，インターネット利用率の比較的多い層が，そのままアンケートの回答者背景に反映されていると推察できる．

本アンケートの内容のなかから，ここでは，河川景観満足度に影響をあたえる項目(説明変数)として，

- 水量が富に流れている(水量).
- 自然がしっかり保全され，残っている(自然度).
- 水質がしっかり保全されている(水質).
- ゴミがなくてきれい(ゴミ).
- 時の変化(季節・気象・朝夕)がすばらしい(四季).
- 水辺の遊びなど，人々の活動がたのしい(親水性).
- 安全で安心する(安全性).
- 鳥や魚や蛍など生物に親しめる(生態系).
- 堰や橋などが美しい(橋・堰).
- コンクリートが少なくてよい(コンクリート).

をとりあげ，全体の満足度を表す項目(応答変数)として，「対象となる河川に満足していますか」をとり上げる．ワード研究所のアンケートでは，その他の項目として，4項目「水辺の町並みや田園と調和して美しい」「日々の生活の中で潤いを与えてくれる」「整然として統一感がある」「全体としての雰囲気がすばらしい」がある．ただし，これらの項目は，河川に対する印象を意味しており，河川景観満足度に対する具体的な意味として解釈できない．そのため，本論文では，これらの項目は除外した．

表-1にアンケート項目と応答での相関行列を示す．自然度，水質そして四季の順で応答との相関が高かった．他方，水量との相関が最も低かったものの，相関係数が0.342であり，さほど低くなかった．水質とゴミ，水質と自然度の相関係数が0.6以上だった．

(1) 重回帰分析の適用

既存のアンケート結果の分析に倣い，重回帰分析を実施した．このときの回帰係数の一覧を**図-2**に示す．その結果，自然度の回帰係数が最も高く，次いで橋・堰が高かった．他方，水量，水質，生態系といった項目の回帰係数が低かった．因に，重回帰分析の適用では，多重共線性の回避あるいはモデルの簡便化(安定性)意図して，変数選択を行うことがある．本データに対してもCp統計量に基づく変数選択を実行したが，取捨される変数はなかった．このとき，全ての回帰係数に対するt検定のp値は，有意水準0.0001のもとで有意であり，係数の適切性が示唆されている．

さらに，自由度調整済み寄与率 $R^{2*}=0.580$ であることから，比較的良くあてはまっている．ただし，橋・堰の相関係数が2番目に低かったにもかかわらず，重回帰係数が自然度に次いで高かった．さらに，自然度や四季といった景観のポジティブな要因が，河川景観満足度に強く寄与している一方で，ゴミや水質といったネガティブな要因の効果がそれほど高くなかった．

(2) CART法の適用

図-3にCART法での結果を示す．ここで，灰色の四角の上側の数字がふし内の応答の平均値を表しており，下側の数字がふしに含まれる標本サイズを表している．全観測値は，先ず，自然度によって分割され，その点数が2以下の観測値は次いで橋・堰によって分割され，それ以外の観測値はゴミによって分割され，最終的に9個の終結ふしに分割された．このとき，分割変数として自然度，橋・堰，水質，ゴミ，四季，および親水性が選択された．他方，生態系や水量，安全性，そしてコンクリートといった項目は出現しなかった．したがって，重回帰分析でも回帰係数が最も高かった自然度による影響が最

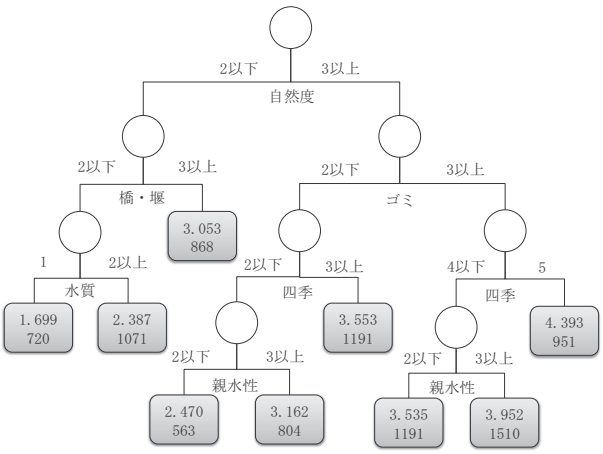


図-3 CART 樹木の結果

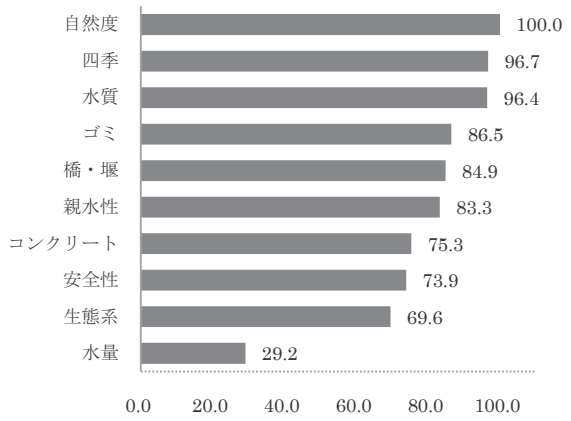


図-4 CART 法の変数重要度

も顕著であった．また，重回帰係数の低かった水質によって分岐されるふしも存在した．

このときの変数重要度を**図-4**に示す．自然度での重要度が最も高く，次いで四季，水質，ゴミの順で高かった．すなわち，河川に対してポジティブな質問項目に対する重要度が最も高かった．次いで，ネガティブな質問項目が続いた．他方，水量や生物など，川辺からは視覚的に捉えにくい要因に対する重要度は高くなかった．多くの傾向が重回帰解析での結果に類似したものの，重回帰係数の低かった，水質の重要度が比較的高かった．水質は応答だけでなく，ゴミおよび自然度とあいだの正の相関関係が認められた．そのため，重回帰分析では多重共線性が生じた惧れがある．これに対して，CART法では，多重共線性を回避できるため，水質の重要度が増加したと考えられる．このときの寄与率を残差平方和の10重クロスバリデーション推定値に基づいて推定した結果は，$R^2_{cv}=0.479$ であり，重回帰分析を下回った．応答と説明変数の相関係数より，これらの線形関係の強さが示唆されており，これにより，ステップ関数で近似するCART樹木でのあてはまりが悪かったと考えられる．

(3) MART法の適用

MART法では，損失関数(4.2)として，最小2乗損失，最小絶対損失，Huberの損失関数などが用いられるが，ここでは，他の方法との比較のため最小2乗損失を用いた．MART法は，ステイジワイズ戦略のなかで，複数の樹木を構成し(アンサンブル学習させていき)，その加法形によってモデルを推定する．**図-5**にブースティング回数と最小2乗損失のプロット示す．ここでは，損失関数の交差確認推定量と学習標本での損失関数の推定値を示す．学習標本では，ブースティング回数を増加させるほど損失が減少する傾向にある．これに対して交差確認推定量は，あるブースティング回数からは飽和傾向を示した．交差確認推定量の結果，本データにおける最適ブースティング回数は549回であることが示唆された．このときの変数重要度のプロットを**図-6**に示す．CART樹木と同様に自然度の重要度が最も高く，次いで四季，水質，ゴミの順で変数重要度が高かった．すなわち，ポジティブ・イメージの要因の影響が最も高く，ネガティブ・イメージの要因がそれに続く傾向にあった．これは，MART樹木のステイジワイズ戦略では，最初の樹木(1回目のブースティング樹木)の変数重要度の影響が最も強いためであると推察される．しかしながら，MART法の変数重要度は自然度，および四季の変数重要度の高さが際立っていた．このときの寄与率をCART法と同様の流儀で計算したところ，$R^2_{cv}=0.655$ と最も高い値を示し

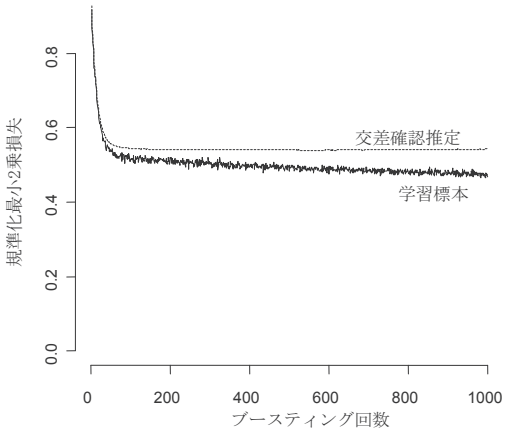


図-5 ブースティング回数と規準化 2 乗損失のプロット

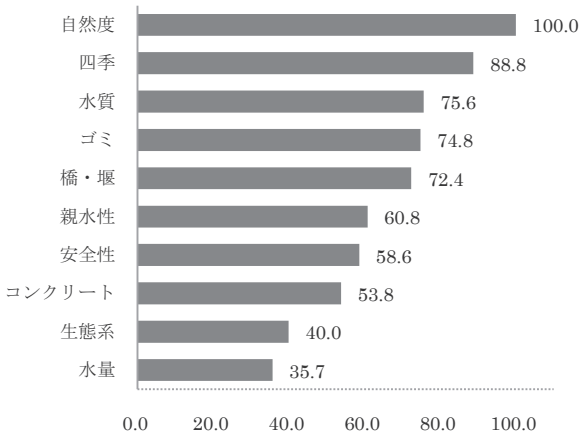


図-6 MART 法の変数重要度

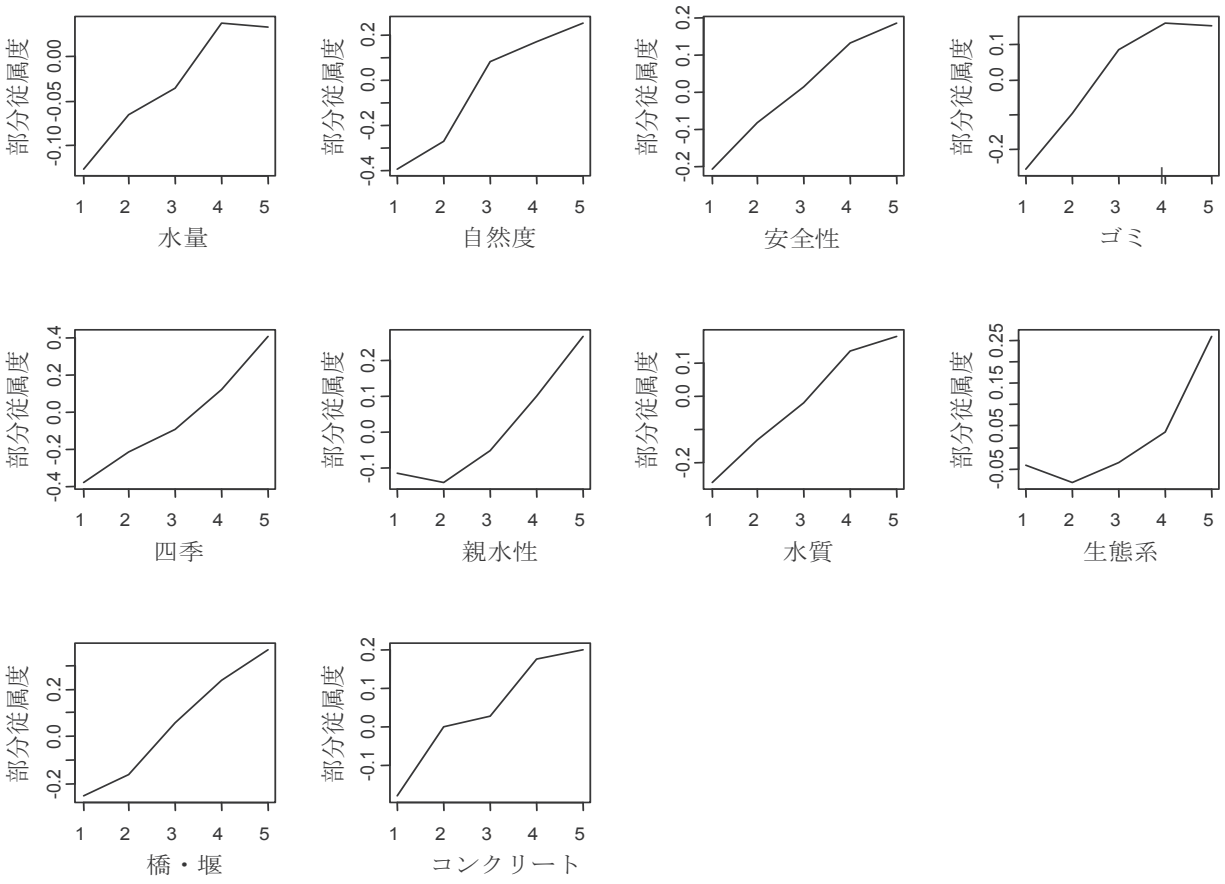


図-7 MART 法に対する 1 次部分従属度プロット
(1:不満である，2：やや不満である，3：どちらでもない，4：やや満足である，5：満足である)

た．すなわち，寄与率の最も低かったCART樹木をアンサンブルさせることで，大幅に寄与率を上昇させることができた．

次いで，部分従属度を省察することで，各要因が応答に与える影響を評価した(**図-7**)．四季，自然度，安全性といった要因は，おおよそ直線傾向を示していることから，線形構造をもつことが示唆された．水質，ゴミ，水量は「満足している」と「やや満足している」でほぼ影響には変化がなかったものの，「どちらでもない」から減少傾向を示した．さらに，生態系では，「満足している」と答えた被験者以外にはほぼ影響がなかった．これらの要因は非線形傾向を示しており，重回帰分析ではこれらの傾向を捉えることができなかった．

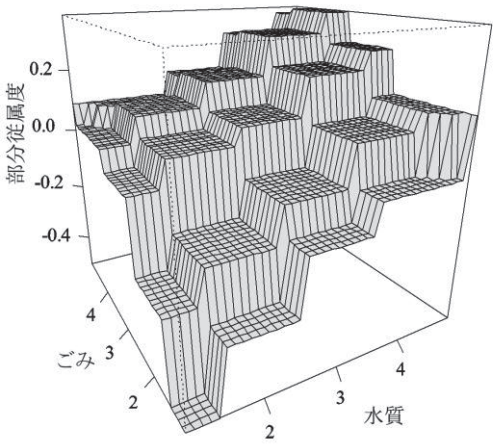
すなわち，CART法およびMART法では，水質での相対的重要度が高いという結果が得られたが，重回帰分析では水質の回帰係数が低く，応答に対する影響の低さが示唆された．これは，重回帰分析では，交互作用効果を自動的に取り込むことができなかったためであると推察される．

表-1では，水質とごみのあいだに高い相関関係が示唆された．そのため，ごみと水質の河川景観満足度に対する2次交互作用を省察するために，これらの2次部分従属プロットを描いた(**図-8(a)**)．その結果，2次交互作用が示唆される傾向を得ることができなかった．すなわち，これらの変数間での交互作用はみられなかった．

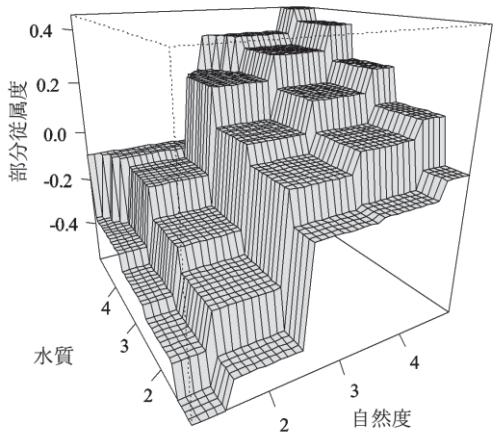
水質と自然度の相関係数も高かった．また，CART法の結果(**図-3**)では，自然度と水質の交互作用が示唆された．水質と自然度の2次部分従属プロットを**図-8(b)**に示す．自然度が「不満である」と答えた被験者は，水質に「満足している」と答えた場合でも，部分従属度が殆ど上がることはなかった．これに対して，自然度が「どちらでもない」以上の満足度もつ被験者では部分従属度は，急激に上昇し，水質に対して満足度が高くなるにつれ，その傾向が顕著だった．したがって，水質は自然度に不満をもつ被験者には殆ど影響がなく，自然度に満足している被験者にとっては，その影響が顕著だった．

(4) 結果の要約

3種類の方法によって河川景観満足度を解析した．ここでは，MART法が他の手法に比して有用だった点に注



⑷水質・ごみの部分従属プロット



⑹ 水質・自然度の部分従属プロット

図-8 MART 法に対する 2 次部分従属プロット

目しながら要約する.

相対的重要度のプロットより，自然度，すなわち河川周辺の環境が河川そのものの満足度にも強い影響を与えることがわかった．このことは，四季の変化，および堰・橋の変数重要度が高かったことからも推察できる．また，自然度，あるいは四季といったポジティブな要因のほうが，ゴミあるいは水質といったネガティブな要因よりも河川景観満足度に影響を与えることが示唆された.
河川のゴミの有無はその満足度に強い影響を及ぼすものの，水量や生物など，川辺からは捉えにくい変数の重要度は高くなかった.

重回帰分析では，交互作用効果を自動的に摘出できないため，MART法の部分従属プロットで示唆された水質と自然度による交互作用効果を適切に捉えられなかった.他方，MART法およびその部分従属度プロットでは，この交互作用効果に関する有用な示唆を与えることができた．すなわち，自然度に不満がある場合には，水質が河川の満足度に殆ど影響しないことがわかった.

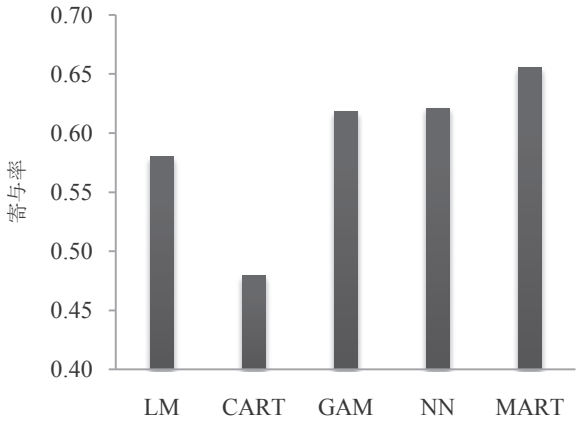


図-9：諸種の回帰手法に対する寄与率
(LM:重回帰分析，CART：分類回帰樹木法，GAM：一般化加法モデル，NN：ニューラルネット，MART：多重加法型回帰樹木)

応答に対して線形関係をもつ数個の説明変数(安全性，水質，四季)が示唆された．そのため，CART法の寄与率が最も低かった．しかしながら，MART法では，CART樹木をアンサンブルさせることにより，その欠点を補うことに成功した.

MARTのモデル適合度を示すために，諸種の回帰手法とMART法の性能を寄与率で比較した結果を図-9に示す.ここではMART法の対照手法として，重回帰分析，一般化加法モデル，ニューラルネット， CART法を用いた.その結果，CART法が最も低い適合度を示した．一般化加法モデルおよびニューラルネットではニューラルネットの方が僅かに良好な適合を示したが，MART法はニューラルネットを上回る適合性能を示した．ニューラルネットでは，モデルの非線形構造を捉える事ができるものの，高次の交互作用を捉えることができない．これに対して，MART法では，回帰器に樹木を用いることで，この問題に対処することができる．このことがモデル適合の好さに反映されたと考えられる.

すなわち，MART法では，非線形構造や交互作用効果に関する多くの知見が得られるだけでなく，高いモデル適合度も得られた.

5．結び

本研究では，河川の満足度に影響を与える要因をアンケート調査の結果に基づいて探索した．アンケート調査の分析には，通常，重回帰分析が用いらるが，これらの要因が単純な線形結合によって結びついていることはなく，要因間の交互作用構造，あるいは応答と要因のあいだの非線形構造を含むことは少なくない．ただし，これらの構造を捉えることは困難である.

本論文では，予測確度と結果の解釈の両方を満たすこ

とができる，多重加法型回帰樹木法(MART法)をアンケート調査の分析に応用し，その有用性を既存の回帰分析手法と比較した．その結果，水質と自然度のあいだに内在する交互作用構造を部分従属プロットにより捉えることができ，そして個々の要因の結果との線形/非線形関係をグラフィカルに提示できた．さらに，MART法は応答に対する説明変数の影響の大きさを変数重要度で評価できる．その結果，河川景観満足度にポジティブな要因(自然度，四季)のほうがネガティブな要因(ゴミ，水質)よりも影響度が高いことが示された．

謝辞：本研究の資料調査において(株)ワード研究所の大島章嘉氏には多大なご協力を頂いた．厚く謝意を表する．

参考文献

1) Witten, I.H., Frank, E : *Data Mining: Practical Machine Learning Tools And Techniques*, Morgan Kaufmann Publishers, Fan Fransisco, 2005
2) 河川景観の形成と保全の考え方検討委員会編：河川景観デザイン，（財）リバーフロント整備センター，2007
3) 建設省中国地方建設局太田川工事事務所：河川構造物の景観デザインマニュアル（案），1990
4) 島谷幸広 編著：河川風景デザイン，山海堂，1994
5) 小路剛志，藤田光一：景観評価指標を用いた都市河川の景観分析, 土木計画学研究講演集, vol.32, 269, 2005
6) 杉本知之，下川敏雄，後藤昌司：樹木構造接近法と最近の発展，計算機統計学, No.18, pp.123-164, 2005
7) 土木学会編：水辺の景観設計，技報堂出版，1988
8) Hastie, T., Tibshirani, R : *Generalized Additive Models*, Chapman & Halls, London, 1990
9) Hastie, T., Tibshirani, R., Friedman. J.H : *The Elements of Statistical Learning: Data mining, inference and prediction*. Springer, NewYork, 2001
10) Friedman, J.H. : Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, Vol.29, pp.1189-1232, 2001
11) Breiman, L. : Bagging procedure, *Machine Learning*, Vol. 26, pp.123-140, 1996
12) Breiman, L.: Using adaptive bagging to debias regression, *Technical Report 547, Statistics Dept. UCB*. 1999
13) Breiman, L. : Random Forests, *Machine Learning*, 45, 5-32, 2001.
14) Breiman, L., Friedman, J.H. Olshen, R.A., Stone. C.J. : *Classification and Regression Trees*. CRC Press, Florida , 1984
15) Freund, Y. , Schapire, R.E. : Experiment with a new boosting algorithm. *Machine Learning: Proceedings of the Thirreen International Conference*, pp. 148-156
16) Ripley, B.D. : *Pattern Recognition and Newral Networks*, Chambridge University Press, London, 1996
17) 市民満足学会・（株）ワード研究所：河川景観満足度調査中間報告書，2006
18) 和田安彦， 道奥康治， 和田有朗：自然環境と河川環境の評価に関する研究,土木学会論文集G, Vol.63, No.3, pp.168-178, 2007

（2008. 10. 7 受付）

ENSENBLE LEARNING METHOD FOR THE RIVER SATISFACTORY QUESTIONNARIE DATA

Toshio SHIMOKAWA, Yukari MUTOH, Taku MISONOU
and Shinichi KITAMURA

In the fields of river landscape study, it is one of important subjects to explore influencing factors on satisfactory of river landscape. The multiple regression analysis is one of the useful statistical tools for exploration of the factors. This multiple regression analysis and related regression analyses (ex. principle regression analysis, partial least square and so on) are usually based on the linear combination of explanatory variables. However, the observations obtained in practice rarely satisfy this constrained model structure. In this paper, we focused on one of the ensemble learning methods, "multiple additive regression tree (MART) method". The performance of MART method was evaluated by the river satisfactory questionnate data[17]. As a result, the MART method gave more prediction performance and provided attractive interpretation than other methods widely used.