Published in final edited form as:

Biometrics. 2014 September ; 70(3): 707–710. doi:10.1111/biom.12187.

Discussion of “Combining Biomarkers to Optimize Patient Treatment Recommendation”

Eric B. Laber*, Anastasios A. Tsiatis, Marie Davidian, and Shannon T. Holloway

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203, U.S.A

1. Introduction

We congratulate the Kang, Janes, and Huang (hereafter KJH) on an interesting and powerful new method for estimating an optimal treatment rule, also referred to as an optimal treatment regime. Their proposed method relies on having a high-quality estimator for the regression of outcome on biomarkers and treatment, which the authors obtain using a novel boosting algorithm. Methods for constructing treatment rules/regimes that rely on outcome models are sometimes called indirect or regression-based methods because the treatment rule is inferred from the outcome model (Barto and Dieterich, 1988).

Regression-based methods are appealing because they can be used to make prognostic predictions as well as treatment recommendations. While it is common practice to use parametric or semiparametric models in regression-based approaches (Robins, 2004; Chakraborty and Moodie, 2013; Laber et al., 2014; Schulte et al., 2014), there is growing interest in using nonparametric methods to avoid model misspecification (Zhao et al., 2011; Moodie et al., 2013). In contrast, direct estimation methods, also known as policy-search methods, try to weaken or eliminate dependence on correct outcome models and instead attempt to search for the best treatment rule within a pre-specified class of rules (Orellana, Rotnitzky, and Robins, 2010; Zhang et al., 2012a,b; Zhao et al., 2012; Zhang et al., 2013). Direct estimation methods make fewer assumptions about the outcome model, which may make them more robust to model misspecification but potentially more variable.

We derive a direct estimation analog to the method of KJH, which we term *value boosting*. The method is based on recasting the problem of estimating an optimal treatment rule as a weighted classification problem (Zhang et al., 2012a; Zhao et al., 2012). We show how the method of KJH can be used with existing policy-search methods to construct a treatment rule that is interpretable, logistically feasible, parsimonious, or otherwise appealing.

2. Setup and Notation

We assume that the available data are (D_i, T_i, Y_i) , $i = 1, \dots, n$, which comprise n independent and identically distributed copies of (D, T, Y) , where $Y \in \mathbb{R}^p$ denotes biomarker information; $T \in \{0, 1\}$ denotes treatment received; and $D \in \mathbb{R}$ denotes the outcome of interest coded so

*laber@stat.ncsu.edu.

that higher values are better, as is customary in the treatment regime literature. To match the development of KJH, we assume that the data are collected in a randomized clinical trial so that $\pi(t|y) = P(T = t|Y = y)$ is known by design, where $\varepsilon < \pi(t|Y) < 1 - \varepsilon$ for some $\varepsilon > 0$ with probability one. This setup includes the binary outcome considered by KJH as a special case with the roles of $D = 1$ and $D = 0$ interchanged.

A treatment rule g is a map from the domain of Y into that of T , so that a patient presenting with $Y = y$ is recommended treatment $g(y)$. Define the value of a rule g , denoted $V(g)$, as the expected outcome if all patients are treated according to g . An optimal regime, g^{opt} , satisfies $V(g^{\text{opt}}) \geq V(g)$ for all other rules g under consideration. Under suitable conditions (e.g., Zhang et al., 2012b) $V(g) = \mathbb{E}[\mathbb{E}\{D|Y, T = g(Y)\}]$. Letting $Q(y, t) = \mathbb{E}(D|Y = y, T = t)$, $g^{\text{opt}}(y) = \arg \max_t Q(y, t)$. The foregoing expression motivates estimating g^{opt} by first estimating the function $Q(y, t)$ by regressing D on Y and T to obtain $\hat{Q}(\hat{y}, t)$ and then obtaining $\hat{g}_{\text{Reg}}(y) = \arg \max_t \hat{Q}(\hat{y}, t)$. We refer to approaches of this form as regression-based methods; the method proposed by KJH is of this type.

An alternative approach is to construct first an estimator for $V(g)$, say $\hat{V}(g)$, and subsequently obtain $\hat{g}_{\text{Val}} = \arg \max_{g \in \mathcal{G}} \hat{V}(g)$ for some suitable class of rules $\mathcal{G}$. We refer to approaches of this form as policy-search methods. Policy-search and regression-based methods are not clearly delineated. For example, one could estimate $V(g)$ by

$n^{-1} \sum_{i=1}^n \hat{Q}\{Y_i, g(Y_i)\}$, in which case $\hat{g}_{\text{Val}} = \hat{g}_{\text{Reg}}$ provided that $\hat{g}_{\text{Reg}} \in \mathcal{G}$. Let $\hat{B}_{\text{KJH}}(Y)$ be the estimator of $Q(Y, 1) - Q(Y, 0)$ proposed by KJH. Because $Q\{Y, g(Y)\} = Q(Y, 0) + g(Y)B(Y)$, the method of KJH can be viewed as a policy-search estimator of the form $\arg \max_{g \in \mathcal{G}} n^{-1} \sum_{i=1}^n g(Y_i) \hat{B}_{\text{KJH}}(Y_i)$. However, the success of this approach relies on the correctness of the outcome model.

Generally, policy-search methods employ estimators for $V(g)$ that do not rely heavily on correctness of an outcome model (Zhang et al., 2012b; Zhao et al., 2012; Zhang et al., 2013). In the clinical trial setting considered here, both the inverse probability weighted estimator (IPWE) and the augmented inverse probability weighted estimator (AIPWE) described by KJH are consistent estimators for $V(g)$, pointwise in g , that do not require a consistent outcome model. We assume subsequently that the AIPWE is used to estimate $V(g)$. Let $\mathbb{P}_n$ denote the empirical measure so that $\mathbb{P}_n f(Z) = n^{-1} \sum_{i=1}^n f(Z_i)$. The AIPWE is

$$\hat{V}_{\text{AIPWE}}(g) = \mathbb{P}_n \left[\frac{1_{T=g(Y)} D}{\pi_c(Y; g)} - \frac{1_{T=g(Y)} - \pi_c(Y; g)}{\pi_c(Y; g)} \hat{Q}\{Y, g(Y)\} \right], \quad (1)$$

where 1_v is one if v is true and zero otherwise, $\pi_c(Y; g) = \pi\{g(Y)|Y\}$, and $\hat{Q}(\hat{y}, t)$ is an estimator for $Q(y, t)$. The AIPWE can be viewed as adding a mean zero but negatively correlated term to the IPWE to gain efficiency (Scharfstein, Rotnitzky, and Robins, 1999). Define

$$\hat{C}(D, Y, T) = \frac{T}{\pi(1|Y)} D - \frac{T - \pi(1|Y)}{\pi(1|Y)} \hat{Q}(Y, 1) - \frac{1 - T}{1 - \pi(1|Y)} D - \frac{T - \pi(1|Y)}{1 - \pi(1|Y)} \hat{Q}(Y, 0),$$

and $\hat{C}_i = \hat{C}(D_i, Y_i, T_i)$. Let $\hat{W}_i = |\hat{C}_i|$ and $\hat{Z}_i = 1_{\hat{C}_i > 0}$. Zhang et al. (2012a) showed that

$$\arg \max_{g \in \mathcal{G}} \hat{V}_{\text{AIPWE}}(g) = \arg \min_{g \in \mathcal{G}} \mathbb{P}_n \hat{W} 1_{\hat{Z} \neq g(Y)},$$

and thus the policy-search objective function using (1) is equivalent to a weighted misclassification error with predictor-label pairs $\{(Y_i, \hat{Z}_i)\}_{i=1}^n$ and weights $\{\hat{W}_i\}_{i=1}^n$. Taking this approach, any classification algorithm that can incorporate weights can be used to estimate an optimal treatment regime. In the next section, we use this framework to develop a policy-search boosting algorithm for estimating g^{opt} .

3. Value Boosting

3.1 Policy-Search Boosting Algorithm

To apply boosting using this classification perspective, we require an algorithm that can accommodate weights. However, standard boosting algorithms assume unweighted data. To provide intuition and to match the development of KJH, we describe a simple boosting algorithm with a logistic regression model as the base classifier, although other choices are possible. To gain intuition, consider first a thought experiment. Suppose that each weight $\hat{W}_i$ is a rational number expressed as q_i/r_i , where q_i and r_i are positive integers. Let N denote the smallest common multiple of $r_1, \dots, r_n$, and create a new, augmented data set with $q_i N/r_i$ copies of $(Y_i, \hat{Z}_i)$. Let $\mathcal{D}_{\text{Aug}} = \{(Y_{ik}, \hat{Z}_{ik})\}_{k=1, \dots, q_i N/r_i, i=1, \dots, n}$ denote this augmented data set, where $Y_{ik} \equiv Y_i$ and $\hat{Z}_{ik} \equiv \hat{Z}_i$ for $i = 1, \dots, n$. In this case

$$\hat{V}_{\text{AIPWE}}(g) \propto \sum_{i=1}^n \sum_{k=1}^{q_i N/r_i} 1_{\hat{Z}_{ik} \neq g(Y_{ik})},$$

which is the (unweighted) misclassification rate of the rule g evaluated on the augmented data set. Thus, we can apply a boosting algorithm without modification to the augmented data set. The algorithm we consider is as follows.

1. Initialize weights $\omega_{ik}^{(1)=1}$, $k = 1, \dots, Nq_i/r_i$, $i = 1, \dots, n$.
2. Repeat for $b = 1, \dots, B$
 - a. Fit a logistic regression model $\hat{g}_n^{(b)}$ to $\mathcal{D}_{\text{Aug}}$ with weights $\omega_{ik}^{(b)}$, $k = 1, \dots, q_i N/r_i$, $i = 1, \dots, n$.

- b.** Compute $m_{ik}^{(b)} = 1_{\hat{g}_n^{(b)}(Y_{ik}) \neq Z_{ik}}$ and weighted misclassification error

$$e^{(b)} = \sum_{i,k} \omega_{ik}^{(b)} m_{ik}^{(b)} / \sum_{i,k} \omega_{ik}^{(b)}$$

- c.** Set $\omega_{iK}^{(b+1)} = \omega_{ik}^{(b)} \{(1 - e^{(b)}) / e^{(b)}\}^{m_{ik}}$.

- 3.** The final boosted treatment rule is

$$\hat{g}_n(Y) = \begin{cases} 1 & \text{if } \sum_{b=1}^B \log \{(1 - e^{(b)}) / e^{(b)}\} \hat{g}_n^{(b)}(Y) > \frac{1}{2} \sum_{b=1}^B \log \{(1 - e^{(b)}) / e^{(b)}\}, \\ 0 & \text{otherwise.} \end{cases}$$

This algorithm is not computationally feasible because $\mathcal{D}_{\text{Aug}}$ is potentially very large.

However, it suggests a simple weighted procedure that can be applied to the original data.

First, note that step 2(a) is equivalent to fitting logistic regression to obtaining $\hat{g}_n^{(b)}$ using the

original data $\{(Y_i, \hat{Z}_i)\}_{i=1}^n$ with weights $\omega_i^{(b)} = \omega_{i1}^{(b)} \hat{W}_i$. Second, defining $m_i^{(b)} = 1_{\hat{g}_n^{(b)}(Y_i) \neq \hat{Z}_i}$, the weighted misclassification at iteration b of the boosting algorithm is

$$e^{(b)} = \frac{\sum_{i,k} \omega_{ik}^{(b)} m_{ik}^{(b)}}{\sum_{i,k} \omega_{ik}^{(b)}} = \frac{\sum_{i=1}^n (N q_i / r_i) \omega_{i1}^{(b)} m_{i1}^{(b)}}{\sum_{i=1}^n (N q_i / r_i) \omega_{i1}^{(b)}} = \frac{\sum_{i=1}^n \hat{W}_i \omega_{i1}^{(b)} m_i^{(b)}}{\sum_{i=1}^n \hat{W}_i \omega_{i1}^{(b)}}.$$

This suggests the following algorithm, which we term value boosting.

- 1.** Initialize weights $\omega_i^{(1)} = n^{-1}$, $i = 1, \dots, n$.
- 2.** Repeat for $b = 1, \dots, B$
 - a.** Fit a logistic regression model $\hat{g}_n^{(b)}$ to $\{(Y_i, \hat{Z}_i)\}_{i=1}^n$ with weights $\omega_i^{(b)}$, $i = 1, \dots, n$.
 - b.** Compute $m_i^{(b)} = 1_{\hat{g}_n^{(b)}(Y_i) \neq \hat{Z}_i}$ and weighted misclassification error
$$e^{(b)} = \sum_i \omega_i^{(b)} \hat{W}_i m_i^{(b)} / \sum_i \omega_i^{(b)} \hat{W}_i$$
 - c.** Set $\omega_i^{(b+1)} = \omega_i^{(b)} \{(1 - e^{(b)}) / e^{(b)}\}^{m_i}$.
- 3.** The final boosted treatment rule is

$$\hat{g}_n(Y) = \begin{cases} 1 & \text{if } \sum_{b=1}^B \log \{(1 - e^{(b)}) / e^{(b)}\} \hat{g}_n^{(b)}(Y) > \frac{1}{2} \sum_{b=1}^B \log \{(1 - e^{(b)}) / e^{(b)}\}, \\ 0 & \text{otherwise.} \end{cases}$$

The form of the above algorithm can be viewed as a version of the adacost algorithm (Fan et al., 1999) and is thus one of many potential boosting algorithms that could be used to construct a direct search analog of the algorithm proposed by KJH.

3.2 Boosting the Value or Boosting the Contrast

Value-boosting focuses on iteratively improving performance of a treatment regime in terms of the estimated value. The method of KJH iteratively attempts to improve an estimator of the contrast function $B(y) = Q(y, 1) - Q(y, 0)$. Because the optimal treatment regime is $g^{\text{opt}}(y) = 1_{B(y) > 0}$ and for any treatment regime, say g , $V(g) - \mathbb{E}Q(Y, 0) = \mathbb{E}g(Y)B(y)$, the problems of estimating the contrast and maximizing the value are closely related. However, value-boosting directly targets g^{opt} rather than B , which may be advantageous in settings where g^{opt} is markedly more parsimonious than $B(y)$; because g^{opt} is a function of B , it is necessarily more parsimonious in some sense. An exaggerated example that illustrates this

point is $Q(y, t) = t \left(\sum_{j=1}^p y_j^2 \right) y_1$, in this case $B(y) = \left(\sum_{j=1}^p y_j^2 \right) y_1$ is a complex nonlinear function of all p variables whereas $g^{\text{opt}}(y) = 1_{y_1 > 0}$ is a univariate linear threshold. However, the operating characteristics of policy-search and regression-based estimators have not been directly compared on examples of this type; we believe such a comparison would provide valuable information on the differences between these two classes of methods.

4. Simulation Experiments

To examine the finite sample performance of the value boosting algorithm, we use the seven simulation scenarios considered by KJH with $n = 500$. We use a logistic working model of the form $Q(y, t) = P(D = 1 | T = t, Y = y) = \text{expit}(\tilde{\gamma}^\top \beta_0 + t \tilde{\gamma}^\top \beta_1)$, where $\tilde{\gamma} = (1, y^\top)^\top$; β_0, β_1 are unknown parameters; and $\text{expit}(u) = e^u / (1 + e^u)$. We estimate $Q(y, t)$ using maximum likelihood. Our base classifier is a logistic model of the form

$P(Z=1 | Y=y) = \text{expit}(\psi_0^\top \tilde{y} + \psi_{12} y_1 y_2 + \psi_{13} y_1 y_3 + \psi_{23} y_2 y_3 + \psi_{11} y_1^2 + \psi_{22} y_2^2 + \psi_{33} y_3^2)$. We estimate the performance of value boosting and the method proposed by KJH using the mean θ value as proposed by KJH computed using a test set of size 10^4 and 1000 Monte Carlo replications. Table 1 shows the estimated performance of each method. Value boosting is comparable with KJH in all scenarios except Scenario 6, where it shows a marked advantage. However, value boosting is more variable than KJH, which is anticipated given that policy-search methods are typically more variable than regression-based methods. The simulations suggest that the boosting method proposed by KJH and value boosting may warrant further investigation.

5. Parsimony, Feasibility, and Interpretability

Boosting and other “black-box” estimation algorithms have a strong track record for predictive performance, especially on benchmark data sets. However, it is not apparent that such methods are suited to inform clinical practice, guidelines, or research. One of the strongest features of policy-search methods is that it is possible to control the class of potential treatment rules. Consequently, one can constrain the estimated treatment rule to be interpretable, low-cost, logistically feasible, and so on. However, boosting these methods to improve value may destroy some or all of the foregoing features.

Consider three approaches for estimating an optimal treatment strategy: (A1) optimize within a small but informative class of regimes; (A2) optimize within a large but difficult to interpret class; and (A3) optimize within a large but difficult class and then project the

estimated regime onto a smaller, interpretable class. Approach (A1) is taken by Orellana et al. (2010), Zhang et al. (2012a,b), and Zhang et al. (2013). In addition to the benefits of interpretability and feasibility imposed on the estimated regime, in some contexts, (A1) is appealing because the best regime within a pre-specified class of regimes is of independent interest. Approach (A2), taken by Zhao et al. (2011), Zhao et al. (2012), and Moodie et al. (2013), uses the predictive power of machine-learning methods and reduces the risk of model misspecification. Approach (A3) attempts to maintain both interpretability and performance (e.g., Breiman and Shang, 1996); furthermore, the “residual” of the estimated treatment regime relative to its projection onto the smaller class may be informative about where and how the smaller class is not sufficiently expressive.

The trade off between flexibility and interpretability is ubiquitous in statistical modeling. However, the importance of each component depends on the context in which the model is estimated. Estimation of an optimal treatment regime is often conducted as a secondary analysis aimed at generating scientific hypotheses and informing follow-up studies. In contrast, KJH seem to be primarily interested in constructing a high-quality treatment regime. Their boosting method seems ideally suited for this purpose.

Acknowledgments

This work was supported by NIH grant P01 CA142538.

References

- Barto, A.; Dieterich, T. Reinforcement learning and its relation to supervised learning. In: Si, J.; Barto, AG.; Powell, WB.; Wunsch, D., editors. Handbook of Learning and Approximate Dynamic Programming. New York: Wiley; 2004. p. 45-63.
- Breiman, L.; Friedman, JH.; Olshen, RA.; Stone, CJ. Classification and Regression Trees. New York: Chapman and Hall/CRC; 1984.
- Breiman, L.; Shang, N. Born again trees. 1996. Available at <ftp://ftp.stat.berkeley.edu/pub/users/breiman/BATrees.ps>
- Chakraborty, B.; Moodie, EEM. Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine. New York: Springer; 2013.
- Fan, W.; Stolfo, SJ.; Zhang, J.; Chan, PK. Adacost: misclassification cost-sensitive boosting. Proceedings of the Sixteenth International Conference on Machine Learning (ICML'99); Bled, Slovenia. June 1999; 1999. p. 97-105.
- Laber E, Linn K, Stefanski L. Interactive model-building for Q-learning. Biometrika. 2014 in press.
- Moodie EEM, Dean N, Sun YR. Q-learning: Flexible learning about useful utilities. Statistics in Biosciences. 2013;1–21.
- Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal treatment regimes, part I: Main content. International Journal of Biostatistics. 2010; 6(2):Article 8. [PubMed: 21969994]
- Robins, JM. Optimal structured nested models for optimal sequential decisions. In: Lin, DY.; Heagerty, PJ., editors. Proceedings of the Second Seattle Symposium on Biostatistics. New York: Springer; 2004. p. 189-326.
- Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. Journal of the American Statistical Association. 1999; 94:1096–1120.
- Schulte P, Tsiatis A, Laber E, Davidian M. Q- and A-learning methods for estimating optimal dynamic treatment regimes. Statistical Science. 2014 in press.

- Zhang B, Tsiatis AA, Davidian M, Zhang M, Laber E. Estimating optimal treatment regimes from a classification perspective. *Stat.* 2012a; 1:103–114. [PubMed: 23645940]
- Zhang B, Tsiatis AA, Laber EB, Davidian M. A robust method for estimating optimal treatment regimes. *Biometrics.* 2012b; 68:1010–1018. [PubMed: 22550953]
- Zhang B, Tsiatis AA, Laber EB, Davidian M. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika.* 2013; 100:681–694.
- Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association.* 2012; 107:1106–1118. [PubMed: 23630406]
- Zhao Y, Zeng D, Socinski MA, Kosorok MR. Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics.* 2011; 67:1422–1433. [PubMed: 21385164]

Table 1

Performance of value boosting and the method proposed by KJH in terms of mean θ based on 1000 Monte Carlo data sets, $n = 500$. Monte Carlo standard deviations are given in parentheses. Scenario indicators correspond to those defined in Section 3.2 of KJH.

Scenario	True θ	Value Boosting	KJH
1	0.127	0.113 (0.007)	0.118 (0.004)
2	0.124	0.125 (0.007)	0.128 (0.004)
3	0.134	0.129 (0.007)	0.134 (0.003)
4	0.066	0.036 (0.016)	0.059 (0.014)
5	0.095	0.072 (0.012)	0.077 (0.007)
6	0.139	0.126 (0.005)	0.045 (0.015)
7	0.142	0.124 (0.011)	0.111 (0.010)