Variable Selection for Optimal Decision Making

Lacey Gunter[1,2], Ji Zhu[1], and Susan Murphy[1,2]

[1] Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA
[2] Institute for Social Research, University of Michigan, Ann Arbor, MI 48109, USA

Abstract. This paper discusses variable selection for medical decision making; in particular decisions regarding which treatment to provide a patient. Current variable selection methods were designed for use in prediction applications. These techniques often leave behind small but important interaction variables that are critical when the goal is decision making rather than prediction. This paper presents a new method designed to find variables that aid in decision making and demonstrates the method on data from a clinical trial for treatment of depression.

1 Variable Selection for Decision making

We consider variable selection in the simplest decision making setting in which one must decide between two actions (usually treatments). Prior to taking an action, we obtain observations about a subject $X = (X_1, X_2, ..., X_p)$, and using this information we choose an action A. We then receive a response, R, an unknown random function based on the action taken and the observations and patient outcomes subsequent to the action. The response, R, gives us some indication of the desirability of the chosen action. A policy, π, is a decision rule mapping the space of observations, X, into the space of the actions, A. The goal is to find a policy π^*, which maximizes the response. Alternate policies are compared via the expected mean response denoted by $V_\pi = E_\pi[R]$. V_π is called the (average) Value for the policy π [1]. The optimal policy, π^*, is defined as

$$\pi^* = \arg\max_\pi V_\pi = \arg\max_\pi E_\pi [R] .$$

A simple example is a clinical trial to test two alternative treatments. The baseline covariates are the observations, the assigned treatment is the action and the response could be the patient's condition post treatment. The goal is to discover which treatment is optimal for any given patient, using the trial data.

Variable selection is often needed in decision making applications. Currently, variable selection for decision making in artificial intelligence is predominantly guided by expert opinion. In clinical trials, the combination of predictive variable selection methods and statistical testing of a few interaction variables suggested by expert opinion are most commonly used [3].

When selecting variables for decisions making, it is useful to distinguish between variables included merely to facilitate estimation as opposed to variables included in the decision rules. *Predictive* variables are variables used to reduce the

variability and increase the accuracy of the estimator. Variables that prescribe the optimal action for a given patient are *prescriptive* variables. For optimal estimation, it is best to select both types of variables. However, only prescriptive variables must be collected when implementing the policy.

For a variable to be *prescriptive*, it must have a qualitative interaction with the action [3]. A variable X_i is said to qualitatively interact with the action, A, if there exists at least two disjoint, non empty sets $S_1, S_2 \subset space(X_j)$ for which

$$\arg\max_a E[R|X_j = x_{1j}, A = a] \neq \arg\max_a E[R|X_j = x_{2j}, A = a],$$

for all $x_{1j} \in S_1$, and $x_{2j} \in S_2$. These variables are useful because they help decipher which action is optimal for each individual patient.

To illustrative this idea, see the plots in Figure 1. Figure 1(a), shows a variable, X_1, that does not interact with the action, A. Figure 1(b) shows a non-qualitative interaction between a variable, X_2, and A. In both plots the optimal action is always $A = 1$. Knowing the value of X_1 or X_2 is useful for predicting R for a given action, but will not effect which action should be chosen. Figure 1(c), shows a qualitative interaction between a variable, X_3, and A. In this plot, the optimal action is $A = 0$, for $X_3 \leq .5$ and $A = 1$ for $X_3 > .5$. Knowing X_3 will impact the choice of action, thus it is useful for decision making.

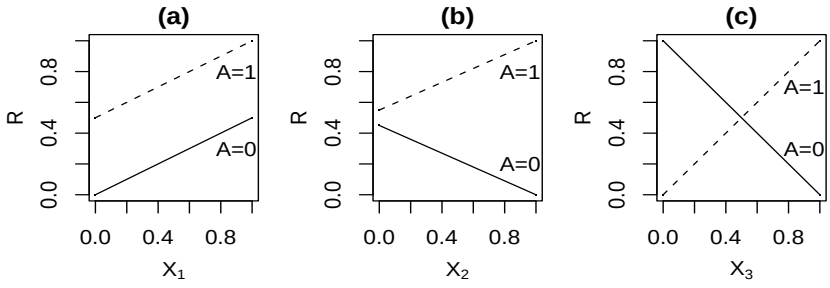

Fig. 1. Plots illustrating interactions. Plot (a) shows no interaction, (b) shows a non-qualitative interaction and (c) shows a qualitative interaction

The degree to which a *prescriptive* variable is useful depends on two factors:

1. *Interaction*: the magnitude of the interaction between the variable and the action. For an action with two possible values, $A \in \{0, 1\}$, this is the degree to which $E[R|X = x, A = 1] - E[R|X = x, A = 0]$ varies as x varies
2. *Proportion*: the proportion of patients whose optimal choice of action changes given a knowledge of the variable. If $a^* = \arg\max_a E[R|A = a]$, this is the proportion of patients for which $\arg\max_a E[R|X = x, A = a] \neq a^*$

Consider the plots in Figure 2. Figure 2(a) shows the relationship between R, A, and a variable X_4, with an underlying plot of the distribution of X_4. Figures 2(b) and 2(c) are similar to 2(a), but for variables X_5 and X_6. Notice that X_4 and X_5 have the same distribution. However, the interaction between X_4 and A is stronger than the interaction between X_5 and A. So the effect of choosing

the optimal action is much greater given X_4 than X_5. Now notice that X_4 and X_6 have the same relationship with R and A, but are distributed differently. The distribution of X_4 is centered at the intersection, so half of the patients would do better choosing $A = 0$ over $A = 1$. Whereas, the proportion of patients benefiting from choosing $A = 0$ is much smaller with X_6.

In the next section we present a new method that ranks the variables in X based on their potential for a qualitative interaction with A. The method is based upon the *interaction* and *proportion* factors for qualitative interactions.

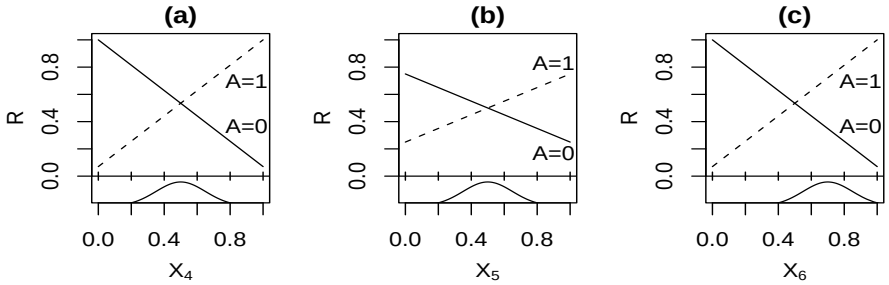

Fig. 2. Plots of qualitative interaction usefulness factors. Plot (a) shows large interaction and proportion, (b) shows smaller interaction, (c) shows smaller proportion

2 Variable Ranking for Qualitative Interactions

Assume we have a data set of $i = 1, ..., n$ patients, with $j = 1, ..., p$ baseline observations for each patients. Also assume we have an action, $A \in \{0, 1\}$, and a response, R, for each patient. Given an estimator of $E[R|X_j = x_{ij}, A = a]$ say $\hat{E}[R|X_j = x_{ij}, A = a]$, define the following two quantities for $j = 1, ..., p$:

$$D_j = \max_{1 \le i \le n} D_{ij} - \min_{1 \le i \le n} D_{ij}, \quad \text{and}$$

$$P_j = p_{1j}(1 - p_{1j}), \text{ with } p_{1j} = \frac{1}{n} \sum_{i=1}^{n} I\{D_{ij} > 0\},$$

where $D_{ij} = \hat{E}[R|X_j = x_{ij}, A = 1] - \hat{E}[R|X_j = x_{ij}, A = 0]$. D_j is a measure of the magnitude of the interaction. P_j is a measure of the proportion of patients affected by a change in the optimal choice of action due to the inclusion of X_j.

These two quantities can be combined to make a score:

$$U_j = \left(\frac{D_j - \min_{1 \le k \le p} D_k}{\max_{1 \le k \le p} D_k - \min_{1 \le k \le p} D_k} \right) \left(\frac{P_j - \min_{1 \le k \le p} P_k}{\max_{1 \le k \le p} P_k - \min_{1 \le k \le p} P_k} \right).$$

The score U_j can be used to rank the variables in X. It has been defined generally to allow for different models of the relationship between R, X, and A. In the section that follows, we used a linear model for the estimator $\hat{E}$.

Since variable ranking is more of a first pass method and not a final variable selection method, we suggest a full algorithm for variable selection in [4]. The algorithm is briefly summarized below.

1. Select important predictors of R in X using any predictive variable selection method
   (a) Use cross-validation to choose the tuning parameter value that gives the best predictive model
2. Rank the variables in X using U_j and select the top k in rank
   (a) Use the important predictors of R selected in step 1 to decrease the variance in the estimator $\hat{E}$
3. Use any predictive variable selection method to select from the important predictors of R selected in step 1, A, and the k interactions chosen in step 2
   (a) Use cross-validation to choose the tuning parameter value that yields a policy with the highest estimated Value

In the next section we reference this algorithm as New Method U. We used Lasso [2] for our predictive variable selection method. We tested the performance of this new method on a wide range of realistically simulated data. We compared the new method against a single Lasso fit on X, A and all interactions between X and A with the penalty parameter chosen by cross-validation on the prediction error. We reference this comparison method as 'Standard Lasso'. The new method appeared to find qualitative interactions better than the standard Lasso and selected variables resulting in policies with higher average Values. Please see [4] for detailed simulation design and results.

3 Nefazodone CBASP Trial

To demonstrate the use of this method, we applied it to data from a depression study to determine which variables might help *prescribe* the optimal depression treatment for each patient. We also tried the standard Lasso for comparison.

The Nefazodone CBASP trial was conducted to compare the efficacy of three alternate treatments for chronic depression. The study randomized 681 patients with chronic major depressive disorder (MDD) to either Nefazodone, cognitive behavioral-analysis system of psychotherapy (CBASP) or the combination of the two. For detailed study design and primary analysis see [5]. We considered $p = 64$ baseline variables detailed in [4]; these variables made up the X matrix. The outcome, R, was the last observed 24-item Hamilton Rating Scale for Depression score [6], post treatment. For lack of space we only compared Nefazodone against the combination treatment. We had responses from 440 patients in this subset.

We used bootstrap [2] to reduce the variance obtained from the different cross-validation splits as follows. We took 100 bootstrap samples of the data and ran the new method and standard Lasso on each sample. We recorded the interaction variables selected and the sign of their coefficients for each sample.

Figure 3 shows plots of the absolute value of the percentage of time an interaction was selected with a positive coefficient minus the percentage of time an interaction was selected with a negative coefficient under the standard Lasso and the new method U. We used this percentage adjustment to help reduce the number of spurious selections. The standard Lasso selected a large number of the variables over 20% of the time. The new method selected only 3 variables

a substantial percentage of the time. These variables were 2 indicators dealing with *alcohol* and a *somatic anxiety* score. For detailed results see [4].

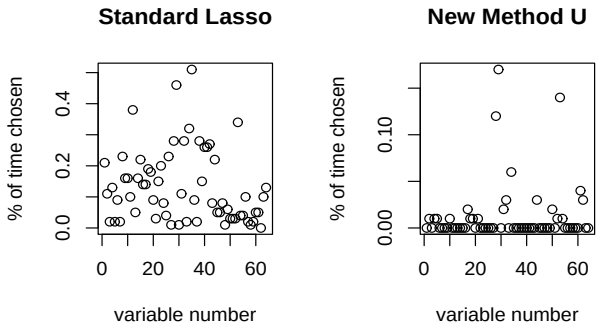

Fig. 3. Variable selection results for Nefazodone CBASP. In each plot, the x-axis is the variable number, the y-axis is the adjusted percent of times variables were selected.

4 Conclusion

In this paper, we discussed when a variable is important in decision making and why variable selection techniques designed for prediction may not perform well in a decision making setting. We presented a new technique explicitly designed to select variables for decision making. The method was tested in simulations and we demonstrated it on a depression data set. The results look promising.

Acknowledgments We wish to thank Martin Keller and the investigators of [5] for use of their data, and gratefully acknowledge Bristol-Myers Squibb for helping fund the study. We also acknowledge financial support from NIH grants R21 DA019800, K02 DA15674, P50 DA10075, and NSF grant DMS 0505432 and technical support from A. John Rush, MD, University of Texas Southwestern.

References

1. Sutton R.S., Barto A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA (1998)
2. Hastie, T., Tibshirani, R., Friedman, J.: Elements of Statistical Learning. Springer-Verlag, New York (2001)
3. Gail, M., Simon, R.,: Testing for Qualitative Interactions Between Treatment Effects and Patient Subsets. Biometrics **41** (1985) 361372
4. Gunter, L., Murphy, S.A., Zhu, J.:Variable Selection for Optimal Decision Making. Technical Report, Univsersity of Michigan Statistics Department, Ann Arbor, MI.
5. Keller, M.B., McCullough, J.P., Klein, D.N.et al.: A Comparison of Nefazodone, the Cognitive Behavioral-analysis System of Psychotherapy, and Their Combination for Treatment of Chronic Depression. N. Engl. J. Med. **342** (2000) 331-366
6. Hamilton, M.: Development of a rating scale for primary depressive illness. Br. J. Soc. Clin. Psychol. **6** (1967) 278-96