nature research

https://doi.org/10.1038/s42256-020-00264-0

Supplementary information

Exploring the cloud of variable importance for the set of all good models

In the format provided by the authors and unedited

nature research

https://doi.org/10.1038/s42256-020-00264-0

Supplementary information

Exploring the cloud of variable importance for the set of all good models

In the format provided by the authors and unedited

Supplementary Information

We introduce an additional experiment to complement our main text.

In designing practical classification models, we might desire to include other considerations besides accuracy. For instance, if we know that when the model is deployed, one of the variables may not always be available, we might prefer to choose a model that does not depend as heavily on that variable. For instance, let us say we deploy a model that provides social services to children. In the training set we possess all the variables for all of the observations, but in deployment, the school record may not always be available. In that case, it would be helpful, all else being equal, to have a model that did not rely heavily on school record. It happens fairly often in practice that the sources of some of the variables are not trustworthy or reliable. In this case, we may face the same tradeoff between accuracy and desired model characteristics of the variable importance. This section provides an example where we create a trade-off between accuracy and variable importance; among the set of accurate model, we choose one that places less importance on a chosen variable.

We study a dataset about mobile advertisements documented in [1], which consists of surveys of 752 individuals. In each survey, an individual is asked whether he or she would accept a coupon for a particular venue in different contexts (time of the day, weather, etc.) There are 12,684 data cases within the surveys.

We use a subset of this dataset, and focus on coupons for coffee shops. Acceptance of the coupon is the binary outcome variable, and the binary covariates include zeroCoffee (takes value 1 if the individual never drinks coffee), noUrgentPlace (takes value 1 if the individual has no urgent place to visit when receiving the coupon), sameDirection (takes value 1 if the destination and the coffee shop are in the same direction), expOneDay (takes value 1 if the coupon expires in one day), withFriends (takes value 1 if the individual is driving with friends when receiving the coupon), male (takes value 1 if the individual is a male), and sunny (takes value 1 if it is sunny when the individual receives the coupon).

	upper bound	lower bound	
zeroCoffee	1.31	1.19	more important
noUrgentPlace	1.16	1.06	↑
sameDirection	1.07	1.03	|
expOneDay	1.06	1.00	|
withFriends	1.02	1.00	|
male	1.01	1.00	↓
sunny	1.00	1.00	less important

Table 1: Bounds on model reliance within the Rashomon set. Each number represents a possibly different model. This table shows the range of variable importance among the Rashomon set.

We compute the VIC for the class of logistic regression models. Rather than providing the corresponding VID, we display only coarser information about the bounds of the variable importance in Table 1, sorted by importance.

Obviously, whether a person ever drinks coffee is a crucial variable for predicting if she will use a coupon for a coffee shop. Whether the person has an urgent place to go, whether the coffee shop is in the same direction as the destination, and whether the coupon is going to expire immediately are important variables for prediction too. The other variables seem to be of minimal importance.

	Least Reliance on noUrgentPlace		Least Error Logistic Model	
	β	VI	β	VI
intercept	0.57		0.07	
zeroCoffee	-2.18	1.27	-2.03	1.26
noUrgentPlace	**0.70**	**1.06**	**1.05**	**1.10**
sameDirection	-1.30	1.05	-0.93	1.04
expOneDay	0.71	1.04	0.64	1.04
withFriends	0.46	1.02	0.17	1.00
male	0.49	1.01	0.18	1.00
sunny	0.24	1.00	0.14	1.00
	logistic loss = 2366		logistic loss = 2296	

Table 2: Reliance and coefficient of the optimal model and the logistic regression estimator. This table shows that as model reliance on noUrgentPlace becomes small, model reliance for all other variables increases.

Suppose we think a priori that the variable noUrgentPlace is unreliable since the survey does not actually place people in an "urgent" situation. in that case, we may want to find an accurate predictive model with the least possible reliance on this variable. This is possible with VIC. Table 2 illustrates the trade-off.

The first and third columns in the table are the coefficient vectors for the two different models. The first column represents the model with the least reliance on noUrgentPlace within the VIC. The coefficients in the third column are for the plain logistic regression model. The second and fourth columns in the table are the model reliance vectors for the two model. The second column is the vector in VIC that minimizes the reliance on noUrgentPlace. The fourth column is the model reliance vector for the plain logistic regression model. By comparing the second and fourth column, we see that we can find an accurate model that relies on noUrgentPlace less. However, its logistic loss is 2366, which is about 3% higher than the logistic regression model. This illustrates the trade-off between reliance and accuracy. By comparing these two columns, we also find that as we switch to a model with the least reliance on noUrgentPlace, the reliance on zeroCoffee, sameDirection and withFriends increases.

References

[1] T. Wang, C. Rudin, F. Doshi-Velez, Y. Liu, E. Klampfl, and P. MacNeille, "A bayesian framework for learning rule sets for interpretable classification," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 2357–2393, 2017.