Proceedings of the 2019 Pre-ICIS SIGDSA Symposium

Special Interest Group on Decision Support and Analytics (SIGDSA)

Winter 12-2019

Predictive Analysis and Out-of-Sample Generalizability of Construct-based Models

Nicholas Patrick Danks

Soumya Ray

Galit Shmueli

Predictive Analysis and Out-of-Sample Generalizability of Construct-based Models

Extended Abstract

Nicholas Patrick Danks
Institute of Service Science,
National Tsing Hua University
nicholasdanks@iss.nthu.edu.tw

Soumya Ray
Institute of Service Science,
National Tsing Hua University
soumya.ray@iss.nthu.edu.tw

Galit Shmueli
Institute of Service Science,
National Tsing Hua University
galit.shmueli@iss.nthu.edu.tw

Introduction

Methodological research in Partial Least Squares Path Modeling (PLS-PM), a construct-based modeling technique, has seen a flurry of efforts to introduce predictive analytic methods. However, there is still confusion about how prediction can be applied to refine theory and integrate with this traditionally inferential technique. We feel that predictive approaches give PLS-PM an opportunity to address the various and growing calls for reform in inferential studies: moving away from categorical validations such as significant versus non-significant (Amrhein et al., 2019); adopting mixed-method approaches that include qualitative methods (Venkatesh et al., 2013); and shifting our singular focus on model fit to other descriptions of model utility (Hair et al., 2019).

We believe that predictive methods offer a promising direction for addressing many of these concerns and calls for change. Prediction describes model utility in terms of generalizability to new and unseen data (Gregor, 2006), which makes it more relevant to practical management decision making (Shmueli et al., 2019). Predictive methods also care about case-level analysis (Shmueli et al., 2016), allowing integration with qualitative approaches to identify and understand highly relevant individual cases.

We propose a mixed-methods, explanatory-predictive framework that: (1) gauges predictive performance of focal constructs, (2) identifies individual cases that might exacerbate overfit and the structural relations that consequently may not generalize well out-of-sample, and (3) guides qualitative analysis to explore the deeper reasons for such conflicts. We demonstrate the practical utility of our proposed analytical framework on a typical technology adoption model in a new context, to show how out-of-sample limitations can be identified with an eye to discovering new insight, and then subsequently reported.

Predicting with Constructs

Researchers building construct-based models are primarily interested in domain-relevant focal concepts known as constructs. Constructs are not directly observable and are thus measured by one or more manifest variables. Given the primacy of constructs in theory building, one major advancement in this study is to devise ways of assessing the prediction of theorized constructs rather than mere measurement items.

But if predictive methods are to be applied to constructs, we require construct scores to be observable for both training and validation data. The determinability of construct scores has made composites the measurement model of choice for recent prediction-oriented literature of construct-based models (Shmueli et al., 2016). We continue in this tradition by calling for prediction-oriented analysis to use composite measurement models. Though we demonstrate our framework using PLS-PM, other newer composite modelling techniques should be equally viable.

The Value of Reconciling Fit and Prediction

Our proposed method seeks to extract value by first quantifying the trade-off between inferential and predictive modelling approaches. As models become better fit to one sample, they become worse in predicting outcomes of other samples (Hastie et al., 2013). Thus, enhancing a model's fit to a training dataset can lead to a model that is *overfit* to that dataset. Overfit, then, is where fit meets prediction and the two intertwine. We expect that understanding and evaluating the causes and ramifications of overfit can offer insights to theory-driven research that is concerned both with explaining phenomena by fitting models and also the practical implications of how that model generalizes to new or future cases.

Proposed Framework

We propose operationalizing constructs as composites and using their scores to evaluate predictive qualities. Datasets of manifest items do not contain construct scores. But we can use an entire sample of data to estimate population parameters and then generate least biased and least inconsistent estimates of actual composite scores. Such estimated actual scores will then allow us to make consistent comparisons of predictive error across subsets of the data. Estimated actual scores can also help us calculate fit and predictive error, which we can summarize for each outcome construct as a mean squared error (MSE) value. We suggest using Leave-one-out Cross Validation (LOOCV) to obtain least biased estimates of predictive error (Burman, 1989) and reporting cross-validated MSE of validation sets from which we can generate aggregate predictive error.

Simply taking the ratio of the difference between fit and predictive error over fit error alone gives a highly interpretable overfit metric that describes the accuracy penalty a model would pay when predicting with data it has not previously seen. However, we caution against arbitrary thresholds of acceptable overfit, as the acceptability of overfit varies across domains. Instead, we suggest that researchers simply report these metrics until such time as meta-analysis can reveal domain- and context-specific guidelines.

We further propose to use the case-level nature of prediction to identify individual cases to which the model is overfitting. We are particularly interested in cases that fit reasonably well when included in the training set, but are hard to predict when our model has not specifically trained for them. We call such cases *predictive deviants*, drawing inspiration from the notion of positive deviance (Zeitlin et al., 1990) wherein exceptional behavior can indicate the need for deeper case study. Likewise, we expect that predictive deviants can inform us of the boundaries of our theories and offer explanations for novel behaviors.

Researchers are often most interested in hypothesized structural paths in their models. Thus, we are especially interested in identifying which structural paths lose generalizability in out-of-sample prediction. To identify which fluctuating paths are affected by predictive deviants, we must first estimate the model including all available cases, and then successively without each of the predictively deviant cases. These fluctuating paths contribute most to overfit and can give us insight into the boundaries of our theories. Researchers can question theory or define boundary conditions for parameters that fluctuate out-of-sample.

Having identified these predictive deviants and the fluctuating paths most affected by them, we can turn to qualitative methods to analyze possible causes. Qualitative investigation allows us to more deeply analyze predictive deviants from their qualitative feedback, and potentially re-engage with them for further information. Such feedback might highlight the need to rethink theory and phenomena, offer practical implications, and point the way for future research. We therefore strongly encourage researchers to collect sufficient qualitative data in their surveys that would allow for such investigation. Additionally, we would strongly caution researchers against throwing away predictive deviant cases from their data without fully understanding the reasons for their impact on the model.

Empirical Demonstration

We conducted a demonstration of the analysis our framework offers on a major portion of the well-regarded UTAUT2 model of system adoption and use (Venkatesh et al., 2012) in the context of project management software. We collected a sample of 216 usable responses from an online survey deployed on Amazon Mechanical Turk (AMT). We conducted our analysis using the SEMinR (Ray et al., 2019) package for R and our own bespoke routines in R for our framework, which we intend to open-source.

We calculated fit MSE at 0.435 and predictive MSE at 0.480, giving us an overfit ratio of 10.3%. Our predictive analysis highlighted five predictive deviants that uncover interesting domain and context relevant phenomena affecting specific relationships in the model. From these, we were able to highlight more nuanced practical and theoretical implications than from our inferential estimation alone.

Discussion

We believe our framework provides practical tools and a process of analysis that allows researchers to extract theoretical value from the predictive performance of models. We also demonstrate how researchers can more readily expand theory, and find the boundaries of their theoretical assumptions and propositions.

To our knowledge, there is no prior framework that outlines which aspects of predictive validity can be meaningfully and practically evaluated to benefit theory-driven composite models. Our proposed procedures are to simply report the predictive penalty paid when moving from model fitting to prediction as a ratio of in-vs-out of sample accuracies, and instead focus investigation on the idiosyncrasies of the data and model that are producing overfit. We emphasize overfit because it describes the shortcomings of the fitting procedure that limits the generalizability of models to new or future data. But rather than categorically testing or modifying the model, we aim to highlight the deviant cases and fluctuating relationships where our inferential models do not generalize well in practical, out-of-sample ways.

Predictive methods have great potential to complement inferential construct-based modeling, while directly addressing the focal constructs, and their hypothesized relationships, that empirical researchers care most about. We have created a combined mixed-methods, inferential-predictive framework that provides a novel analytic tool for researchers, while paying heed to the wider concerns of the scientific community regarding the limitations of extant inferential methods. Armed with this new framework, we hope researchers can explore the boundaries of their theories and draw out more nuanced theoretical and practical implications.

References

Amrhein, V., Greenland, S., and MsShane, B. 2019. “Retire Statistical Significance”. *Nature* (567), pp. 305–307 (https://doi.org/10.1038/d41586-019-00857-9).

Burman, P. 1989. “A Comparative Study of Ordinary Cross-Validation, V-fold Cross-Validation and the Repeated Learning-Testing Methods”. *Biometrika* (76:3), pp. 503-514 (https://doi.org/10.1093/biomet/76.3.503).

Gregor, S. 2006. “The Nature of Theory in Information Systems”. *MIS Quarterly* (30:3), pp. 611-642 (https://doi.org/10.2307/25148742).

Hair, J. F., Sarstedt, M., and Ringle, C. M. 2019. “Rethinking Some of the Rethinking of Partial Least Squares”. *European Journal of Marketing* (53:4), pp. 566-584 (https://doi.org/10.1108/EJM-10-2018-0665).

Hastie, T., Tibshirani, R., and Friedman, J.H., 2013. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (second ed. 2009. Corr. 10th printing 2013 edition). Springer, New York, NY.

Ray, S., Danks, N.P., and Velasquez Estrada, J.M. 2019. seminr: Domain-Specific Language for Building PLS Structural Equation Models. R package version 0.1.0.

Shmueli, G., Ray, S., Estrada, J. M. V., and Chatla, S. B. 2016. “The Elephant in the Room: Predictive Performance of PLS Models”. *Journal of Business Research* (69:10), pp. 4552-4564 (https://doi.org/10.1016/j.jbusres.2016.03.049).

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J. H., Ting, H., Vaithilingam, S., and Ringle, C. M. 2019. “Predictive Model Assessment in PLS-SEM: Guidelines for Using Plspredict”. *European Journal of Marketing* (Ahead of print), (https://doi.org/10.1108/EJM-02-2019-0189).

Venkatesh, V., Thong, J. Y., and Xu, X. 2012. “Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology”. *MIS Quarterly* (36:1), pp. 157-178 (https://doi.org/10.2307/41410412).

Venkatesh, V., Brown, S. A., and Bala, H. 2013. “Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems”. *MIS Quarterly* (37:1), pp. 21-54 (https://doi.org/10.25300/MISQ/2013/37.1.02).

Zeitlin, M. F., Ghassemi, H., Mansour, M., Levine, R. A., Dillanneva, M., Carballo, M., and Sockalingam, S. 1990. *Positive Deviance in Child Nutrition: With Emphasis on Psychosocial and Behavioural Aspects and Implications for Development*. United Nations University: Tokyo.