International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-4 Issue-3, February 2015

Evaluating Prediction Factor Prominence in Academic Domain Selection using Dominance Analysis – Ministry of Higher Education (MoHE), Ibri CAS, Sultanate of Oman

S. Sharmi, Ishtiaque Mahmood, Jehad Bani-Younis

Abstract: - This paper, advocates on a broader use of relative prominence keys as an appendage to multiple regression analysis. The goal of such analysis is to screen the variance among multiple predictors to realize the role played by each predictor in a regression equation. Dominance Analysis is a method to evaluate the relative prominence of the prognosticators. Regrettably, when predictors are correlated, they totally trust on metrics which are flawed indicators of variable importance. Furthermore, the key benefits of two relative prominence analyses, dominance analysis and relative weight analysis, over estimates produced by multiple regression analysis. Here, this investigation helps us to evaluate the importance of the prediction factors involved in determining the criteria's for domain selection of the students. A mockup study was conducted to evaluate the performance of the proposed actions and develop commendations.

Index Terms— Predictor prominence, weight analysis, Dominance Analysis (DA), Multiple Linear Regression (MLR).

I. INTRODUCTION TO RELATIVE PREDICTOR PROMINENCE

A key issue when designing a selection system is how to determine which selection tool may contribute the most to the prediction of the outcome or criterion of interest. Researcher LeBreton states that "Relative prominence refers to the stable contribution each predictor makes to R^2 considering both its individual effect and its effect when combined with the other variables in a regression equation" (LeBreton, 2007). Determining the relative prominence is essential in at most of all circumstances for maximizing the selection system utility. Relative importance is not necessary when a single predictor is used because there are not two variables to relate to one another but organizations generally choose to use a multiple predictor system. There are 4 major strategies to determine relative predictor prominence: (1) regression coefficients, (2) incremental validity, (3) dominance analysis, and (4) relative weight analysis.

Manuscript published on 28 February 2015.
* Correspondence Author (s)
Ms. S. Sharmi, Department of IT, Ibri College of Applied Sciences, Sultanate of Oman.
Ms. Ishtiaque Mahmood, Department of IT, Ibri College of Applied Sciences, Sultanate of Oman.
Dr. Jehad Bani-Younis, Department of IT, Ibri College of Applied Sciences, Sultanate of Oman.

When estimating a selection system for forecaster function, it is more necessary to use standardized regression coefficients so it is possible to know which predictors provide stronger relative importance. Organizations may weigh these predictors with more relative importance in order to improve a selection system's ability to identify applicants with high performance potential. Conversely, attempting to relate unstandardized regression coefficients would be hard and confusing because the variables would be compared on a diverse metric. Hence, any weighting with unstandardized coefficients would also be misleading, because scores would exhibit different values (e.g. 1-10 score range on one predictor versus 1-5 score range on another predictor). In sum, if a regression coefficients strategy is used to determine relative importance of multiple predictors, standardized regression weights should be used.

While the logical and methodological simplicity of this strategy may be considered useful, the strategy does possess an inherent weakness. In his seminal work, Hoffman (1960) claimed that his conception of relative importance measured "independent contribution of each predictor". Other researchers called into question Hoffman's assertion of "independent" contribution of each predictor because it inferred that every other predictor's influence is held constant in the model (Ward, 1962). Hoffman (1962) was forced to reply that his conception of relative weights did not measure independent contribution in that sense. Essentially, if predictors are uncorrelated or irrelevant, standardized regression coefficients equal zero-order correlations, and if the squared regression coefficients are summed, they equal R2.However, when predictors are correlated (as they almost always are in selection systems), a change in one predictor variable will almost assuredly result in a change in all other correlated predictors, and summing squared regression coefficients will no longer equal R2, making decomposition of the effects difficult or near impossible (LeBreton, 2007). That is, the estimates of importance from regression coefficients use total effects, but ignore partial and direct effects, masking the effects of correlated predictors (LeBreton, Ployhart, & Ladd, 2004). Consequently, indices of predictor importance that do not consider the existing relationship between all of these different variables can be misleading (Johnson & LeBreton, 2004).

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

II. ISSUES ADDRESSED ON PREDICTOR SELECTION

The normalization of coefficients using a fractional standard deviation pedals the correlations in other predictors that has been suggested, but this approach still does not consider the partial effects (Bring, 1994). Thus, it is common to report these as indices of relative predictor importance, but it may be unwise to rely solely on them because they yield highly divergent results compared to more recently developed strategies have demonstrated more accuracy in Monte Carlo simulations (Johnson & LeBreton, 2004; LeBreton, et al.,2004). Specifically, because regression coefficients control for, rather than expose partial effects, they can create confusion around which one in a set of correlated predictors is best, making relative ranking ineffective.

A. Correlation Coefficients:

The second major strategy, which is frequently used in conjunction with the first strategy, is the examination of incremental validity in the form of correlations (Azen & Budescu, 2003). This strategy, like the first strategy, is also simple and straightforward in its logic because if one variable is statistically related to another (i.e. criterion variable), then logically it would be important. Essentially, this index gives an indication of how much more of the variance in a criterion variable can be attributed to the inclusion of the selected predictor variable (e.g. increase in R2). However, there are also inherent weaknesses involved in this strategy. A major weakness is that similar to regression coefficients, correlation coefficients do not consider partial and total effects of relationships between predictors and criterion variables (LeBreton, et al., 2004). Essentially, correlation coefficients cannot partition the variance shared between multiple predictors that should be attributed to each predictor. Consequently, squared correlations will only sum to the model's R2. If the predictors are all uncorrelated (LeBreton, 2007). Additionally, model order entry can affect predictor importance, which can mask a predictor's relative importance. For example, if X2 is entered after X1 and they are both highly correlated, X1 will look more important, but if the order is reversed then X2 will look more important. Weaknesses Associated with Both Strategies. Even though the first two approaches are the most commonly used in research and in practice through selection system design, both share weaknesses that warrant consideration (Azen & Budescu, 2003). In academics, the responses of the students on the various prediction factors not only helps us to evaluate the ability of the students but helps to classify their interest on selective major (Bennink Margot et.al, 2014).

B. Why not MLR?

The individual students choice of domain specialization selection may fall within the same cluster (e.g., Software, Data Management, Network, and Security in IT) are likely to be more analogous to each other than to those in another cluster, it is incongruous to use multiple linear regression to analyze such data due to the violation of the assumption of independence of observations. The domain selector identifier number for the domains Software, Data Management, Network, and Security are 1,2,3,and 4 respectively. When data are layered, it is more appropriate to use HLM rather than multiple linear regression. Hierarchical Linear Modeling (HLM) has become a very common analysis approach to determine the relative importance of predictors. As the instance with multiple linear regression, the connotation of individual predictors can be readily tested in HLM to specify each predictor's statistical prominence while governing all other predictors. Though, the standardized coefficient of a predictor as well as its statistical worth might change depending on the subset of predictor variables incorporated in the model.

III. PROPOSED METHODOLOGY

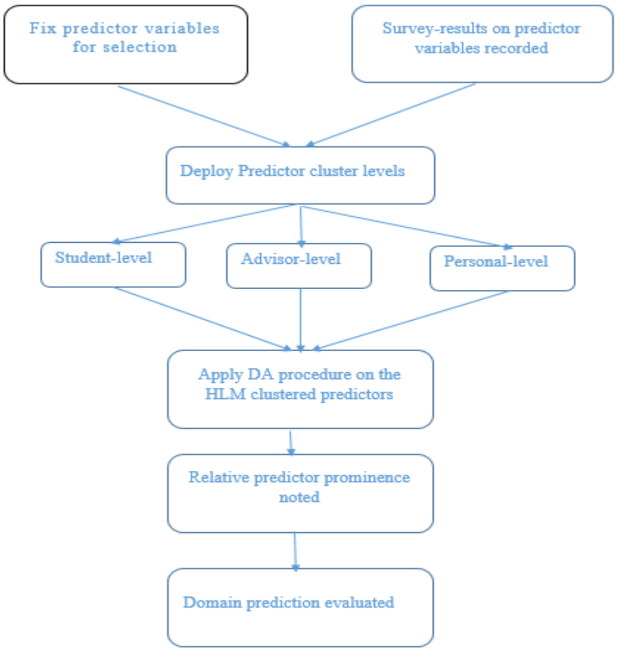

Fig. 1: Proposed frame-work diagram for relative predictor prominence in domain selection.

A. Predictor Clusters:

To predict the student's choice of domain the outcomes from **student-level predictors** are being cast-off (such as the student's gender, creative thinking, logical thinking ability etc.) as well as **advisor-level predictors** (such as the percentage of student's performance in the courses etc.). Or may want to predict outcomes over several points in time (such as compliance with prescribed student activity each term) from time-level predictors (such as the person's choice of workshop attendance, schedule, etc.) as well as **person-level predictors** (such as general attitude towards selecting courses, grades, etc.).To estimate the performance of the various measures commendations are made as to which measures might be optimum for evaluating predictor importance using Dominance Analysis.

B. Dominance Analysis:

In the case of MLR, R^2 is a measure of the descriptive authority of a model. It is defined as the proportion of the total unevenness in the outcome (or response variable (Y_i) as elaborated in equation (1) that is accounted) by the model. For example, given a regression model for n observations ($i = 1\ldots n$) and p predictors, and the variance is measured as shown equation (2).

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \varepsilon_i \quad \text{- Equation (1)}$$

$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} \quad \text{- Equation (2)}$$

Where SST is the total sum of squares and represents the overall variability of the responses, Y_i, while SSE is the error sum of squares and represents the variability of the random error components, ϵ_i. The SSE is thus the portion of the total variability in the response that is unaccounted for by the model (i.e., the predictors $X_1 \ldots X_p$). One method used to determine the importance of a predictor variable in the case of multiple linear regression is DA (Azen & Budescu, 2003). DA rank orders the predictors in footings of prominence. The influence of a predictor is measured with the increase in variance, or the increase in R^2. The impact factor of the predictor X_3 to the subset model consisting of the predictors X_1 and X_2 is the increase in R^2 that occurs when X_3 is added to that subset model; X_3 in this instance the variance between R^2 of the model consisting of X_1, X_2 and X_3 and the R^2 of the model consisting of X_1 and X_2. The prominence of a predictor variable to a given archetypal is thus measured as the increase in the prototype's ability to describe the erraticism of the outcome variable when that predictor is added to the subset model. Comparisons of added aids can be made among the forecasters as they are added to the same subset model, and the authority of one predictor over another is established as the added contribution is superior to the other forecaster's contribution over all subset models. The measure of the added contribution of X_4 to the subset model consisting of X_1 and X_2, and then compare the added assistances of X_3 and X_4 to this subset model. If X_3 adds more than X_4, then X_3 is relatively more vital than X_4 in the context of this subset model. To launch dominance, this kind of association should be made across all subset models.

IV. EMPIRICAL RESULTS ON DA

The strength of this proposed approach detects and identifies the suppressor variables as the Dominance analysis table shows negative dominance index instead of being masked. It also performs a constrained DA by complementing the best predictor variable in the model. This indeed benefits this selection system and allows the forecaster to yield better predictive estimations and enhance its prominence. The predictor variables tagged to the 3 levels of predictor clusters are listed below in Fig 2.

PV.Tags	Predictor-Variables(PV)
X_1	Problem Solving
X_2	Virtualization
X_3	Visualization
X_4	Communication
X_5	Team work
X_6	Dexterity
X_7	Critical thinking
X_8	Field Work
X_9	Technical writing
X_{10}	Creativity

	Coefficients	Std Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	19.47797	3.322141	5.863076	0.02788	5.183947	33.77199	5.183947	33.77199
Visualization	-2.11864	0.809737	-2.61646	0.120281	-5.60266	1.365372	-5.60266	1.365372
Communication	-1.02712	0.259671	-3.95547	0.058375	-2.14439	0.090154	-2.14439	0.090154
Team work	-0.86441	0.31509	-2.74336	0.111153	-2.22013	0.491317	-2.22013	0.491317
Dexterity	-1.23051	0.286195	-4.29955	0.050067	-2.46191	0.000889	-2.46191	0.000889
Critical thinking	1.433898	0.234561	6.113125	0.025731	0.424666	2.443131	0.424666	2.443131
Field Work	-1.99322	0.458417	-4.34805	0.049037	-3.96563	-0.02081	-3.96563	-0.02081

Fig. 2: Predictor variables in 3 different predictive clusters.

The predictive prominence is shown among the following predictor variables after DA. The PV's are Virtualization, Visualization, Communication, Team work, Dexterity, Critical thinking, Field Work.

V.SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.994726
R Square	0.98948
Adjusted R Square	0.952659
Standard Error	0.247016
Observations	90

General Dominance : 0.6583 Field Work	0.5544 Team work	0.7089 Critical thinking	0.6861 Comm	0.5306 Visual	0.6778 Dexterity
Re-Scaled Dominance : 16.3923	15.2955	17.6511	17.0840	14.7007	16.8765

VI. PROBABILITY OUTPUT

Percentile	*Domain Selector id no*
5	1
15	1
25	1
35	1
45	2
55	3
65	3
75	3
85	3
95	4

Fig. 3: Summary & Probability output relative predictor prominence

The overall required accuracy is determined by R^2 and adjusted R^2. R^2 represents the percentage of variance of the output variable (domain selector), which is explained by the variance of the input predictor variable.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The R^2 is 0.989 and adjusted R^2 is 0.952 as shown in Fig 3. The R^2 increases as a new predictor valuable is added to the variable (never decreases), whereas the adjusted R^2 compares and improves the value of the model if and only if the new predictor enhances the chance or otherwise it decreases the value as the expected chance gets reduced. Predicting overall domain selection among the students at various predictor levels using DA has improved the selection of predictor's prominence as shown in the probability output of Fig 3. The relative importance/dominance of the predictors' critical thinking & communication have first level of priority on predicting the domain. The second level of importance to dexterity, field work predictors and the third level of importance to teamwork and visualization variables as depicted in fig 3, the above shaded table of dominance.

VII. CONCLUSION

Despite of the use of regression coefficient and correlation coefficients to determine the predictor prominence, the proposed method of using the DA does point to predictor prominence more accurately. Henceforth, the usage of DA as an appendage gauge is recommended by several researchers to enrich the predictive prominence in future.

REFERENCES

[1] Azen, Razia, and David V. Budescu. "The dominance analysis approach for comparing predictors in multiple regression." Psychological methods 8.2 (2003): 129.
[2] Bennink, Margot, et al. "Measuring Student Ability, Classifying Schools, and Detecting Item Bias at School Level, Based on Student-Level Dichotomous Items." Journal of Educational and Behavioral Statistics 39.3 (2014): 180-202.
[3] Bring, Johan. "How to standardize regression coefficients." The American Statistician 48.3 (1994): 209-213.
[4] Hoffman, Paul J. "The paramorphic representation of clinical judgment." Psychological bulletin 57.2 (1960): 116.
[5] Johnson, Jeff W., and James M. LeBreton. "History and use of relative importance indices in organizational research." Organizational Research Methods 7.3 (2004): 238-257.
[6] Lebreton, James M., Robert E. Ployhart, and Robert T. Ladd. "A Monte Carlo comparison of relative importance methodologies." Organizational Research Methods 7.3 (2004): 258-282.
[7] LeBreton, James M., et al. "A multidimensional approach for evaluating variables in organizational research and practice." Personnel Psychology 60.2 (2007): 475-498.
[8] LeBreton, James M., and Scott Tonidandel. "Multivariate relative importance: extending relative weight analysis to multivariate criterion spaces." Journal of Applied Psychology 93.2 (2008): 329.
[9] Ward Jr, Joe H. "Multiple linear regression models." Computer applications in the behavioral sciences (H. Borko, ed.). Englewood Cliffs, New Jersey: Prentice-Hall (1962).

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)