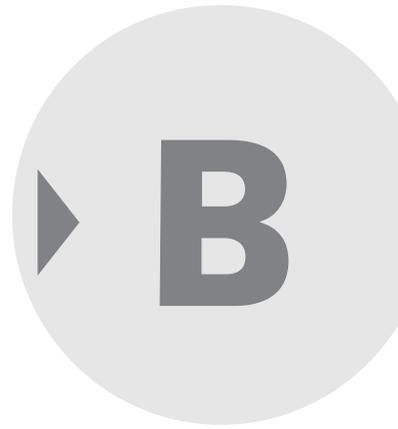


Appendix

LASSO Methodology



LASSO: An Algorithm for Automated Brand Self-Evaluation

Overview

In order to allow users to decide whether their own brand could be further extended, an algorithm was developed to determine a brand's extension based on the LASSO scoring framework. This algorithm allows users to self-evaluate their brand according to the guidelines provided in this book, and receive recommendations as to how optimally the brand is being extended. Using state-of-the-art statistical techniques, the algorithm aims to simulate an expert assessment of brand extension in automated manner, allowing both consistent and objective brand evaluation.

To develop an algorithm that accurately characterizes complex phenomena, the most effective current methods rely on fitting a statistical model to a verified, known set of training examples in a process known as "supervised learning." For the purpose of developing such an algorithm to characterize brand extension, a "gold-standard" dataset of brand evaluations was generated by an expert panel of three brand specialists, and this was used to optimize, train, and evaluate the model. The resulting algorithm

produced by this analysis performs both accurately and consistently, providing a robust solution with which users may evaluate their own brands.

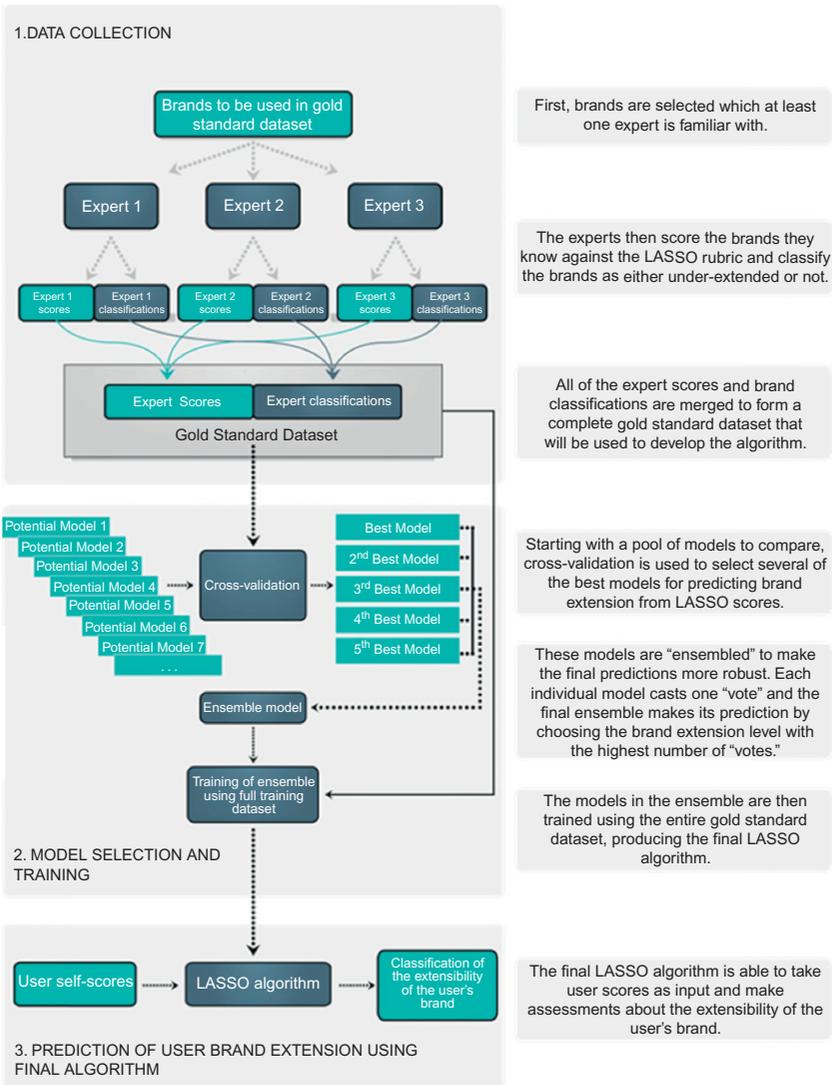
Data Collection

The dataset generated by the expert panel consists of both the LASSO scores and a corresponding determination of brand extension for 56 brands, including brands as famous and large as Coca-Cola, Mickey Mouse, and the NFL and as different as Chupa Chups, FIFA World Cup, Nerf, and *World of Warriors*. The brands that were evaluated and each expert's determination of brand extension (as either under-extended or optimally/over-extended) are listed in Table 8.1. Roughly half of the brands characterized in this dataset were under-extended and the other half were either optimally extended or over extended. Note that this group of brands was selected by the panel to include companies across a diverse range of industries. By including brands of companies both large and small across many industries in this training dataset, the algorithm is able to generalize effectively to characterize a wide spectrum of brands. Indeed, the inclusivity of this training dataset should enable this algorithm to accurately classify brand extension.

To further improve the accuracy and real-world relevance of the algorithm, a subset of 25 of the brands was independently rated by each of the three experts. This overlap allows the model to capture the intrinsic, yet entirely valid, variation in these metrics. In addition, the overlapping set of examples allows a direct comparison of the agreement between predictions made by the algorithm and those made by human experts. See below a process flow chart depicting the steps taken from Data Collection through Model Selection and Training (see next page).

Model Selection

The task of determining whether a brand is extended to an optimal degree or not is best suited to the group of statistical models that aim to classify examples into one category or another, a



process known as “binary classification.” Many binary classification models exist, each with different strengths and weaknesses for various types of datasets and variables. To identify the best model for the problem of classifying brand extension, several of the most powerful model families from conventional statistics and modern machine learning were evaluated and compared.

According to the best practices in statistics, a model cannot be evaluated against the data that it was trained on; an example used to “fit” the model cannot be used to judge the model’s performance, or serious biases will invalidate the results. There are many ways to avoid this bias, and all generally involve splitting the entire dataset into both “training” and “testing” subsets. Here, the model being evaluated is fit on the “training” data, and then predictions are made on the “testing” dataset. The accuracy of these predictions is then used to measure the performance of the model. One of the most effective methods for generating these training and testing datasets is a technique known as “cross validation.” The main benefit of this technique, over other techniques for validating a model, lies in the fact that it evaluates the model on every single example in the original dataset. Because of this, a model that classifies some types of brands much better than others will always be penalized to the same extent, while other methods of model evaluation may rate this model higher or lower in a fairly random manner.

Using this cross-validation framework, five models were chosen that showed promise in predicting the expert classification of a particular brand’s extension. However, to further enhance the accuracy and reliability of predictions made by this algorithm, one additional step was added. Rather than choosing the single best of the five top-performing models to generate the final predictions, the predictions of all five were combined with a machine learning technique known as “ensembling.” Essentially, this technique generates individual predictions for each model, and each model then casts a “vote” for its prediction; these votes are then tallied and the prediction given by a majority of the models is used as the final prediction. For example, if three models predict that a brand is under-extended while the other two predict it is optimally extended, the final “ensembled” model will predict that the brand is under-extended. The power in this technique arises from the fact that the individual models, although performing fairly similar to each other, make mistakes that are not identical. Because the models do not make exactly the same predictions, the majority consensus will more often be right than any individual model. After applying ensembling to the five best-performing models identified with

cross-validation, the algorithm's performance increased significantly to nearly the same level as human experts, as detailed below.

Agreement between Experts

The 25 common brands, for which the brand specialists independently scored on the LASSO rubric and assessed brand extension, provide an observation of the true, inherent variability in brand assessment. While the LASSO framework provides a powerful, quantitative approach to brand assessment, variability is present in all real-world datasets and this must be considered both when generating a model and when evaluating it. While training a model, including this inherent variability actually improves the performance of the resulting model. And, when evaluating the model, the agreement between human experts sets an upper limit on the predictive capabilities that can be expected of such an algorithm. To quantify this variability, the standard deviation between the expert scores for each of the LASSO metrics was determined for each brand in this set (Figure B1). These were then averaged for each brand, providing a look at the inherent ambiguity or complexity in rating each brand (gray bars in Figure B1), as well as for each metric, providing a comparison of the variability for each of the LASSO variables (rightmost set of bars in Figure B1).

Overall, the panel of brand experts showed a very high level of agreement in their LASSO scoring metrics. The overwhelming majority of scores deviated by at most 1 point in only one of the three experts' scores (on a scale of 1–5), suggesting that the LASSO rubric, when properly deployed, is capable of precisely and quantitatively characterizing brands. With regard to the classification of brands as under-extended or not, the expert panel produced a unanimous classification for 19 of the 25 brands, suggesting that it is straightforward to determine brand extensibility in roughly 80% of cases, while one out of every five cases may be more involved and require further consideration. As a matter of reference in Figure B1, 0.47 is the standard deviation for a score where one expert disagrees from the other two by exactly 1 (e.g., expert scores of 4, 4, and 5).

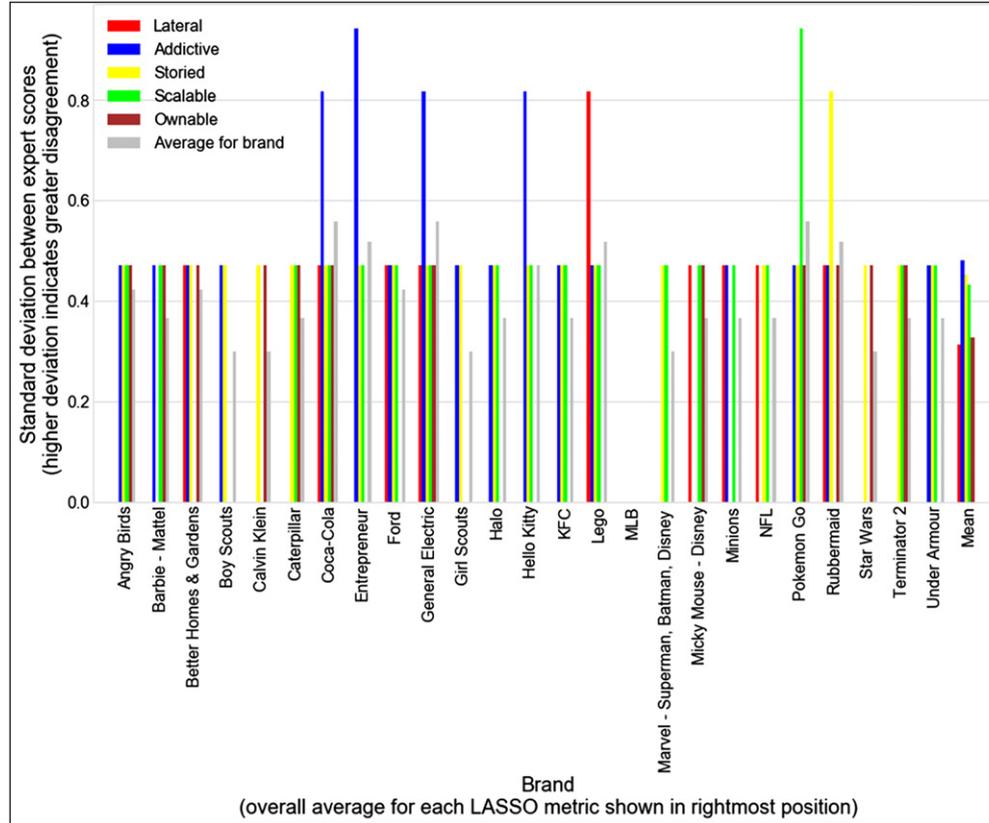


Figure B1. Agreement between Experts on LASSO Scores for Brands Scored by All Experts.

The standard deviation for MLB was 0.00 as indicated in the figure. Please note that the experts first rated brand extension on a more fine-grained five-point scale that was later down-sampled to a simpler “under or not-under extended” rubric. This was done due to limitations with the size of the dataset available, and may have resulted in the experts placing brands in the “slightly-under extended” category with differing frequency. Thus, a coarser rubric may have resulted in slightly higher unanimity between expert classifications.

Algorithm Evaluation

The final algorithm, using an ensemble of five well-performing models, was evaluated using two methods. As both methods involve cross-validation, which is difficult to apply to the scores for more than one expert at a time, only the scores and classifications for one expert (Expert 1) were used to generate the model and predictions for this step. For both methods of algorithm evaluation, cross-validation was used to first predict the extensibility of each brand in the dataset. The first method aims to assess the absolute capabilities of the algorithm to model this dataset, while the second compares the model’s performance with that of the human expert brand specialists.

For the first evaluation, these predictions were compared to the “true” classifications chosen by Expert 1. In this test, the algorithm correctly predicted the brand extensibility for 39 of the 49 total brands (79.6%) assessed by Expert 1. Notably, three of the five best-performing individual models, all of which were used in the final ensemble, correctly predicted the classification of 36 of the 49 (73.6%) brands when evaluated by individually. Although this does not seem to be drastically different from the success rate for the complete ensembled algorithm, the fact that all three top-performing models are able to predict exactly the same number of brand assessments correctly suggests that this may be an upper limit to accuracy of individual models with these data, and ensembling or other such techniques may be required.

The second method of evaluation used the set of brands scored by all three experts to determine how the algorithm's predictions compare to a human's predictions. By comparing the number of times that all three experts agreed on a classification of a brand's extension to the number of times the algorithm correctly predicted the classifications of Expert 1, it is possible to characterize how well the algorithm performs with both ambiguous cases and for brands with more well-defined brand extension. When expert judgment could not consistently determine a brand's extension, the model performed poorly, correctly predicting only 3 of 6 (50%) of the classifications of Expert 1, a result no better than random. When all three experts agree on the brand's extension, however, the algorithm correctly classifies 15 of 19 brands (78.9%), indicating that the algorithm is truly capturing the intricacies of the LASSO scores that impact a brand's extensibility. Last, with this evaluation we are able to directly compare how well Experts 2 and 3 agreed with, or "predicted", Expert 1's classification of these brands with how well the algorithm predicted these classifications. In total, the two other experts predict Expert 1's classification for 19 of these 25 example brands (76%), while the model predicts Expert 1's choices on 18 of 25 examples (72%).

Again, with such a relatively small set of data it is difficult to make detailed inferences from these results, but the results do suggest that the algorithm performs respectably, even when compared to expert brand specialists. This is expected, as the algorithm is trained on data generated by these very experts. Also, since we believe that it is capturing the information relating to brand extension contained in the LASSO metrics, it follows that a larger volume of expert training data in the future will allow the model to better represent this information and become increasingly robust. Additional enhancements such as including industry information in the model and the previously mentioned fine-grained categories for brand extension are likely to further boost the model's accuracy and precision. In summary the algorithm, as it currently exists, provides a repeatable, widely deployable, and inherently objective method for both expert and amateur owners to evaluate their brands.