MPRA

Munich Personal RePEc Archive

Modeling Economic Choice under Radical Uncertainty: Machine Learning Approaches

Anton Gerunov

Sofia University "St. Kliment Ohridski"

January 2016

Modeling Economic Choice under Radical Uncertainty: Machine Learning Approaches

Anton Gerunov[1]

Abstract

This paper utilizes a novel data on consumer choice under uncertainty, obtained in a laboratory experiment in order to gain substantive knowledge of individual decision-making and to test the best modeling strategy. We compare the performance of logistic regression, discriminant analysis, naïve Bayes classifier, neural network, decision tree, and Random Forest (RF) to discover that the RF model robustly registers the highest classification accuracy. This model also reveals that apart from demographic and situational factors, consumer choice is highly dependent on social network effects.

Key words: choice, decision-making, social network, machine learning

Introduction

Understanding economic choice is of huge theoretical and practical importance. On the theoretical front, more intimate understanding of the agents' motivations and actions leads to improved knowledge of economic and social systems and the resulting ability to improve welfare. On the practical side this means better modeling for the benefit of forecasting, marketing efforts, and a variety of other business and public policy needs. How real people make decisions is therefore a natural focal point of interdisciplinary research in business, economics, sociology, psychology and numerous adjacent fields.

For the purposes of modeling, a few simple yet robust statistical techniques are commonly in use that provide insight into choice. Those largely linear methods are suited to small samples of well-defined consumption choices and have seen only very limited improvement in classification and forecasting accuracy in complex environments. With the advent of increased data availability and the explosion of sample sizes, older approaches are beginning to see limits to their usefulness. Novel methods hailing from the growing field of machine learning seem to be better equipped to handle consumption modeling in large, diverse, complex datasets which exhibit marked non-linearity.

We aim to present some of the most popular machine learning algorithms and test them on novel experimental data on consumption choice under uncertainty. Results are then juxtaposed against linear methods such as logistic regression and discriminant analysis to outline the main differences. One is compelled to notice that even with a relatively small sample size, some ensemble models of non-linear classifiers tend to outperform more traditional models. The article is structured as follows. The second section presents classical approaches to modeling choice and popular methods and algorithms from the field of

---

[1] Assistant Professor at Sofia University "St. Kliment Ohridski", Department of Economics.
E-mail: gerunov@uni-sofia.bg

machine learning. The third section presents the experimental data and results, and the fourth discusses major findings. The last section concludes.

Models of Economic Choice

Understanding decision-making in economic context has historically been tightly connected to a view of agent rationality, whereby individuals try to maximize their benefit (or utility), given some objective constraints like budget, production, or endowment. The pinnacle in this view is reached in von Neumann and Morgenstern's theories in the mid-twentieth century (von Neumann & Morgenstern, 1944). They developed an axiomatic theory of choice, whereby consumer had strictly defined preferences in mathematical form (utility function), which they seek to maximize. In such context modeling decisions consists largely of relatively straightforward solutions of optimization problems. A major issue with this approach is that consumers do not act in such a trivial and well-defined way in reality, and thus a number of paradoxes and inconsistencies with theory soon surfaced (Allais, 1953; Ellsberg, 1961).

This school of thought would lead to modeling individual choices by defining the respective agent's preferences in a narrow mathematical sense and then solving a constrained optimization problem given their constraints. Thus total utility $u(x_i)$ is a function of a number of i features x_i. Optimal choice is then derived as the maximum utility, given a set of constraints, S:

$$\max_{S} u = u(x_i), i = 1 \dots n. \tag{1}$$

Research in behavioral economics quickly revealed that this model of economic choice makes up in elegance what it lacks in efficiency. Kahneman and Tversky's (1972, 1979) research outlined a few clear heuristics and biases that people fall prey to when making decisions. Later, Thaler (1980) famously argued for the need for a positive theory that is able to describe behavior in an empirically realistic manner. While experimentation continues to this day with ample evidence for the rich and sophisticated set of factors influencing decisions (Camerer, 2003; Duffy, 2008), a comprehensive mathematical representation remains as of yet elusive. This has spurred practitioners to look for alternative empirical models to understand choice. The key modeling questions is thus the probability of certain (discrete) choice $P(y)$, given other factors x_i. This conditional probability is here denoted as $P(y|x_i)$.

Models were historically selected for their ease of computation and flexibility for representing choice. In the case of discrete choice or classification problems, the **logistic regression** is a popular choice and was pioneered early in modeling problems (Cox, 1958; McFadden, 1981). In this case the probability of choosing a certain outcome is approximated by the logistic function, or:

$$P(y|x_i) = \frac{\exp(\beta_0 + \Sigma\beta_i x_i)}{1 + \exp(\beta_0 + \Sigma\beta_i x_i)}. \tag{2}$$

The estimated beta coefficients show the strength of association between a given independent variable like a demographic or a situational factor, and the dependent one – the

choice. Those coefficients have the interpretation of increasing the odds of selection. The simple regression can be expanded to a multinomial logistics regression and has been widely used in modeling applications (Hyman and Young, 2001; Akinci et al., 2007).

An alternative but still very popular approach is the **linear discriminant analysis**. It aims to classify a binary dependent variable (choice) by constructing the best linear combination of the observed variables (Ripley, 1996). Let us assume the two conditional distributions of the outcome y and predictor x to be: $p(y|x)$ and $p(x|y)$, and further that these two follow a normal distribution with means μ_y and μ_x, and with covariations of σ_{yx} and σ_{xy}, respectively. If the condition $\sigma_{yx} = \sigma_{xy}$ holds, then classification can be obtained via the following condition:

$$\left(x - \mu_y\right)^T \sigma_{yx}^{-1} \left(x - \mu_y\right) - (x - \mu_x)^T \sigma_{xy}^{-1} (x - \mu_x) < T. \tag{3}$$

for some threshold value T. Due to its simplicity and relative ease of interpretation, (linear) discriminant analysis continues to be of use to researchers of consumer choice (Tregear & Ness, 2005; Hansen, 2005).

While traditional modeling has provided many valuable insights into individual decision-making, there are certain limitations to those methods. First of all, they fail to capture some non-linear associations in the data, which can sometimes be of great importance. These can include threshold levels, salient anchors, and tipping points that influence the agent's thinking. Second, they impose a number of stringent assumptions on data at hand such as normality of distribution and deterministic one-way relationships which are not necessarily met in practice. Oftentimes the decision situation and context are subtly nuanced and simple methods fail to capture this complexity.

Third, simple linear methods fail to scale well to large datasets. Traditional indicators of the goodness of a given model such as the p-value tend to be inflated as the number of observations grows, and thus they give a misleading perspective of model fit. All these factors point to the possibility of improving choice classification by introducing novel modeling tools. Those are presented in more detail in the following section and then tested rigorously against experimental data.

Classification problems are a common occurrence from the field of machine learning whereby an algorithm is needed to distinguish between different groups of classes of observations (Hastie et al., 2011; Zhu et al., 2014), e.g. return visitors vs. non-return visitors to a web-site. The parallel to decision making is straight-forward as the different decisions or courses of actions can be easily interpreted as different classes and thus modeled in a simple and intuitive way. The most popular classification algorithms include neural networks, decision trees, ensembles of trees (random forests), Support Vector Machines, and Bayesian approaches such as Naïve Bayes classifiers. While some of them have been making their way into the field of applied modeling and marketing, this process has been somewhat slow. In addition to that, no consensus exists as to what the best method to approach economic choice is.

Naïve Bayes classifiers are applications of the Bayes theorem, whereby the classification problem is solved by constructing the joint probability distributions of variables under interest and then using that for the purposes of class assignment. More specifically, we are interested in the conditional probability distribution of observations y_i over classes C_k given a number of features x_i that are pertinent to the classification problem. Assuming that any feature is independent of the others, this conditional distribution can be described as:

$$p(C_i|x_i) = \frac{1}{Z} p(C_k) \prod_{i=1}^{n} p(x_i|C_k). \tag{4}$$

Here we use Z to denote a scaling factor with $Z = p(x_i)$. Once the conditional probability distribution is algorithmically constructed, the classfier is complete once a decision rule is specified. A common choice is to accept the most likely class, thus assigning an observation y_i to class C_k under the following condition:

$$y_i = \underset{k}{\operatorname{argmax}}\, p(C_k) \prod_{i=1}^{n} p(x_i|C_k). \tag{5}$$

The naïve Bayes classifiers are not as computationally intensive as other machine learning methods but still perform well in practice. While their major assumption of feature independence is rarely achieved in reality and the posterior class probabilities may thus be imprecise, overall classification results are sometimes at par with more sophisticated approaches. A further advantage of the naïve Bayes approach is it tends to scale well, thus making it a plausible modeling option for larger datasets. More details on this approach as well as an overview of its performance over the last decades can be found in Lewis (1998).

Despite its many advantages the naïve Bayes classifier is predominantly used as a specialized classification tool, often alongside other approaches. This is likely due to the fact that alternative algorithms tend to offer better performance and accuracy, and the NB scalability is less of an issue with the decrease in the cost of computational power. Despite this the algorithm has remained in popularity over a long time span, especially in the machine learning context, and there are some applications that use it for modeling consumption choice and sentiment (Ye et al., 2009; Cheung et al., 2003; Huang et al., 2012).

Neural networks are models that are heavily influenced by the way the human brain works. It is structured by neurons that send activation impulses to each other, and so is the overall architecture of the neural network model. The different input nodes (independent variables) send activation impulses in the form of mathematical weighting functions to hidden layers of other nodes, which then produce impulses towards the final output node (dependent variable). Thus each input neuron (independent variable) can affect the class under study through a series of weighted functions, a representation known as the nonlinear weighting sum. If the node is denoted as x, and its activation function is K, the neuron's network function f can be expressed mathematically as:

$$f(x) = K\left(\sum_{i=1}^{n} w_i\, g_i(x)\right). \tag{6}$$

In this case *K* is some pre-defined function, and $g_i(x)$ is a series of functions weighted by w_i. Estimation methods (learning) calculate the optimum weights given a set of conditions such as the functions used and the number of layers. Many neural networks can be trained on the same set of data with varying degrees of complexity. The optimal choice is guided by computational tractability and parsimony. A more detailed description of the features and statistical properties of neural networks can be found in Ripley (1996).

It seems that among the machine learning algorithms, neural networks have found most place in practical problems of modeling consumer behavior. They have been applied and often tested against more traditional models, whereby neural nets tend to produce superior classification results (Gan et al., 2005; Hu et el., 1999; Hu et al., 2008). Despite improved accuracy, there are also a number of disadvantages. Most notably neural networks act as a black box, making interpretation of results very difficult and thus lend themselves to only limited analysis. In addition the learning, calibration and network selection processes can sometimes be daunting to the practitioner.

Support Vector Machines are advanced classification methods, stemming from the field of machine learning (Cortes & Vapnik, 1995). Given two classes of observations they estimate the best classifier by finding the hyperplane in the middle of the largest distance (margin) between the closest points in the two classes. The boundary points are called the support vectors. Essentially, the classification tasks consists of solving quadratic programming problems. More sophisticated SVMs are able to perform accurate non-linear classification by mapping finite dimensional data into higher-dimensional space through a given kernel function and performing the classification task there. We present the simplest of SVMs – the linear one. More concretely, with a set of given points x_i, each of them belonging to class y_i, we can define a hyperplane as follows:

$$\vec{w} \cdot \vec{x} - b = 0 \tag{7}$$

Here $\vec{w}$ is the normal vector to the hyperplane, and *b* is a parameter. The optimization problem aims to find the largest distances between the support vectors, thus minimizing $\|\vec{w}\|$, or:

$$\min\|\vec{w}\|, \tag{8}$$
$$s.t.: \; y_i(\vec{w} \cdot \vec{x} - b) \geq 1$$

The solutions to his problem -, define the optimum classifier for the linear SVM. Extension to the method allow for multiple class classifications, thus making the SVM a versatile tool for modeling discrete choice among multiple alternatives. Research in decision making has already used it to make inference about taste and preference (Bahamonde et al., 2007) and to segment consumer groups and opinions (Li et al., 2006; Xu et al., 2012). While the SVM excels in classification problems, especially under non-linear setting, its popularity is marred by the interpretational difficulty of the model and the iterative process of learning it. This holds particularly true for the selection of appropriate SVM method and its corresponding kernel functions.

Decision trees provide an alternative approach to modeling. Their machine learning operationalization is merely an extension of the classical decision tree model, used in decision science. Given a test data, an algorithm splits classification cases at different nodes by picking the best classifier among the features tested. Let us assume a classification problem with k classes and N_m observations in the region R_m. Thus the probability of observation y_i belonging to class C_k at node m equals:

$$p_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = C_k).$$

(9)

In that case an algorithm splits the classification region to find the best classifier x_i to put a class prediction C_k on observation y_i. Then every observation is classified at node m as the majority class of this node:

$$C(m) = argmax_k(p_{mk}).$$

(10)

More classification nodes are created until some pre-determined number of is reached. After they are so grown, trees can be used to perform classification tasks in validation or test sets. Further details beyond this short overview can be found in Breiman et al. (1984). Decision trees have the great advantage that they are easy to build, yet intuitive to interpret. In their visual form they can also be used for decision-making on the spot, especially in the case of compact tree. Their major drawback stem from the fact that trees tend to overfit data, produce large variation and can be misled by local optima.

Despite this there has been some applications of trees in modeling decisions choices that have produced enlightening results (Lee et al., 2006; Kim et al, 2001). It is sometimes the case that classification and regression trees perform as well as more complex models (Rezi & Athappilly, 2005) while retaining their simplicity and ease of use. In other applications, where overfitting and high variance do tend to be a problem, a common approach to trees is to bag them in an ensemble, the so-called random forest (Breiman, 2001).

Random forests essentially combine a pre-determined number of trees into a large ensemble that can be used for classification or regression. This process initiates by first selecting a bootstrapped subsample of the data from $b = 1$ to B, and then selecting a random number of features to be used for classification. The algorithm then builds a tree in such a way that the best variable/split point is created at node m in the same way as with classification and regression trees. Once the maximum number of trees and their respective number of terminal nodes is reached, those are combined in the forest ensemble $\{T_b\}_1^B$. Just like neural networks, random forests can model both continuous and discrete choice. The rule for continuous choice is:

$$f_k^B = \frac{1}{B} \sum_{b=1}^{B} T_b(y).$$

(11)

Under discrete choice, the classification problem is solved by taking the majority vote of trees as to the class of given observation y_i, thus obtaining:

$$C_k^B(y_i) = majority.vote\{C_b(y_i)\}_1^B. \tag{12}$$

While random forest model are notable for their ease of interpretation and relatively limited parameters, interested readers are directed for more details and nuances to the seminal work by Breiman (2001). Random forests are steadily gaining popularity in consumer modeling and marketing applications but are still not used at a large scale for consumer analytics. Current applications for deriving consumer preference (Bi, 2012) and modeling decisions (Kruppa et al., 2013) hold great promise for their usefulness in further cases and tend to consistently outperform alternative methods.

The wide availability of alternative methods poses the question of the optimal modeling strategy. While this is dependent on the particular dataset and its peculiarities it does seem that some methods tend to perform better than others on a majority of classification problems. For the purposes of testing, we utilize a novel dataset on individual decision making under large uncertainty in a complex environment. This data poses two particular modeling challenges – the possible nonlinearity of relationships and the relatively small sample size. Both of them make this data a suitable and novel test of methods.

Application and Results

Data was collected from a laboratory experiment, performed on 258 young adults (aged mostly 18-28). Under this experimental settings each participant plays a game in which he or she selects a supplier of a fictional good, *omnium bonum*. There are four possible suppliers denoted with letters from A to D (coded as 1 to 4), and each of them makes a quantity offer. After the supplier is selected, the participant is informed of the actual delivery which does not need to coincide with the initial offer – it can be either higher or lower.

Following this information, participants rate their satisfaction with the supplier on a scale from -4 to 4 and then proceed to the next round. Some participants receive real-time information of these ratings on their screens, thus simulating the social effects on consumer choice. Some of the participants receive information on the total production and its growth, while others do not. This is done to emulate the asymmetry of information, characteristic of many decisions in economic context. Participants play a total of 20 rounds in which they make choices in order to maximize the resulting total quantity of the good, which is then monetized at the end of the experiment.

This setting yields a dataset of 5160 discrete economic decisions that can be fed into a classification algorithm. In additional to the age and gender of the consumers (demographics), we also have data on the difference in deliveries in previous rounds (delta1, delta2, delta3), satisfaction in previous rounds which gauges emotional state (satis1, satis2, satis3), previous choices as a proxy for consumption habit (lag1, lag, lag3), time for selection as a measure for expended cognitive effort (tim1), information for production (aggregates), and number of rounds played or experience (round). The endowment effect on consumption is captured by

the variable cumulative results (cum.res) which measures what amount of omnium is already collected as a proxy for individual wealth.

Further, participants are randomly assigned to three different experimental conditions – one of them simulates a growing economy with a constant increase in the amount of the good, and the other – a cyclical economy with rise and fall of production (cycle). The variable social denotes whether players have access to the social network of satisfaction scores. We also construct a measure of reputation of each supplier which is simply the average satisfaction of all the users up to this point (avgrepA, avgrepB, avgrepC, avgrepD).

Summary statistics of the data are shown in Table 1. The key fact from the table is the very large difference between the minimum and the maximum, i.e. the range, of many features. This captures the rich and nuanced context in which agents make decisions – some take a lot of time, and some – very little, with the results producing wide fluctuations in consumer satisfaction. The reputation of suppliers is also very different, with supplier C clearly leading the pack. The experimental design is constructed in such a way that indeed this is the supplier which provides the greatest quantity of the good, given his promise. This design is also reflected in the relative frequency of choice – 33% of players choose supplier C, 28% - supplier B, 21% - supplier D, and 17% - supplier A.

Table 1: Summary Statistics of Experimental Data on Economic Decisions

	Mean	Std. Dev.	Median	Min.	Max.	Range	Skew	Kurtosis	Std. error
gender	0.60	0.49	1.00	0.00	1.00	1.00	-0.39	-1.84	0.01
age	20.67	2.93	20.00	0.00	50.00	50.00	2.84	46.19	0.04
round	10.50	5.77	10.50	1.00	20.00	19.00	0.00	-1.21	0.08
supplier	2.59	1.00	3.00	1.00	4.00	3.00	-0.12	-1.06	0.01
time1	11460.55	13579.18	7040.50	530.00	160343.00	159813.00	3.70	21.20	189.04
satis1	0.94	2.70	2.00	-4.00	4.00	8.00	-0.44	-1.25	0.04
satis2	0.95	2.72	2.00	-4.00	4.00	8.00	-0.44	-1.26	0.04
satis3	0.95	2.70	2.00	-4.00	4.00	8.00	-0.44	-1.26	0.04
delta1	1.55	35.50	0.00	-60.00	60.00	120.00	-0.08	-1.60	0.51
delta2	1.97	35.46	0.00	-55.00	60.00	115.00	-0.07	-1.64	0.52
delta3	1.48	34.53	0.00	-52.00	52.00	104.00	-0.07	-1.66	0.52
cum.res	915.76	554.89	879.00	51.00	2485.00	2434.00	0.20	-1.05	7.72
cycle	0.50	0.50	0.50	0.00	1.00	1.00	0.00	-2.00	0.01
aggregates	0.52	0.50	1.00	0.00	1.00	1.00	-0.06	-2.00	0.01
social	0.49	0.50	0.00	0.00	1.00	1.00	0.03	-2.00	0.01
lag1	2.58	1.01	3.00	1.00	4.00	3.00	-0.11	-1.07	0.01
lag2	2.59	1.00	3.00	1.00	4.00	3.00	-0.11	-1.06	0.01
lag3	2.59	1.01	3.00	1.00	4.00	3.00	-0.11	-1.07	0.02
avgrepA	0.42	1.19	0.29	-3.67	5.00	8.67	0.36	1.69	0.02
avgrepB	1.98	5.57	0.57	-4.33	31.00	35.33	3.51	13.04	0.11
avgrepC	1.46	2.51	0.89	-7.50	11.67	19.17	1.21	3.82	0.05
avgrepD	0.47	1.38	0.33	-3.00	4.50	7.50	0.36	0.63	0.03

The final outcome of the participants are also widely divergent – some have collected as much as 2485 units, while others – much less. All discreet variables can be interpreted as unordered factors, whereas continuous ones – as true numeric values. This dataset can be used both to

test different modeling strategies and to extract substantive knowledge of the consumer choice process.

Estimation of the models presented in the second section follows a straightforward logic. We first divide the dataset in two partitions: one of them, comprising 80% of the total, to be used for training, and the other 20% for testing. The logistic regression, linear discriminant analysis, and the naïve Bayes classifier models are deterministic and easy to calculate from data without additional parameters. Training neural networks showed that the optimal structure would be a net with one hidden layer for both conditions, with the final results obtained from 25 bootstrapped samples.

The classification tree was also built through 25 bootstrapped resamples. In order to avoid overfitting of the ensemble Random Forest, the models across conditions were limited to only 40 trees with 100 terminal nodes each. Data was used to estimate a SVM with linear kernel in both cases with cost = 1, and gamma = 2. Models are learnt on the testing set, and then used to predict classes from the training one.

Results from the two experimental conditions with the training and the test set are presented in Table 2 and Table 3, which outline the classification accuracy of the modeling approaches. When economic decisions are modelled in a situation with no access to social network, algorithms produce up to 62% accuracy in the training, and 40% accuracy in the test set. Classical methods such as discriminant analysis and logistic regression produce relatively good classification rates of 41% in the training set, and about 39% in the testing set. This should be set against the natural benchmark of random classification, and those methods improve significantly upon that.

Most of the tested machine learning approaches do not seem to yield noticeable advantage in terms of accuracy – the correct classifications in the test set for the Naïve Bayes classifier, the Neural Net, the SVM, and the Tree range between 34% and 36%. The relatively low training set accuracies indicate that despite the complexity of methods, there was hardly any model overfitting. The Random Forest classifier registers best results across all seven methods under review and is well ahead of the two main contenders – SVM and the Neural Network.

Table 2: Classification accuracy across methods for conditions without a social network

Method	Training Set		Test Set	
	Point Est.	*95% CI*	*Point Est.*	*95% CI*
Linear Discriminant Analysis	41.2%	(38.9-43.5)	39.1%	(34.9-43.8)
Logistic Regression	41.4%	(39.1-43.6)	38.8%	(34.4-43.5)
Naïve Bayes Classifier	39.4%	(37.1-41.7)	34.8%	(30.3-39.4)
Neural Network	34.2%	(32.1-36.5)	34.3%	(29.9-38.9)
Support Vector Machine	40.4%	(38.1-42.7)	36.3%	(31.9-41.0)
Decision Tree	37.3%	(35.0-39.6)	34.5%	(30.1-39.2)
Random Forest	62.2%	(59.9-64.5)	40.1%	(35.6-44.9)

Consumer choice modeling is more challenging under conditions, where a social network is present. Its inclusion leads to a lowering of successful classification by three to six percentage

points. In this case we observe a deterioration in performance of classical methods and relative improvement for machine learning algorithms. While LDA and logistic regression have almost the same classification ability in the training set of about 40%, they see significant deterioration in the test set. The Neural Network, the SVM and the Tree model perform only slightly better. The Random Forest models achieves the highest accuracy rate both in the training (59%), and in the testing set (37%).

Table 3: Classification accuracy across methods for conditions with a social network

Method	Training Set		Test Set	
	Point Est.	*95% CI*	*Point Est.*	*95% CI*
Linear Discriminant Analysis	40.0%	(37.7-42.4)	33.5%	(29.0-38.2)
Logistic Regression	39.9%	(37.6-42.3)	33.7%	(29.3-38.4)
Naïve Bayes Classifier	37.1%	(34.9-39.5)	31.6%	(27.3-36.3)
Neural Network	34.4%	(32.2-36.7)	34.4%	(29.9-39.1)
Support Vector Machine	38.4%	(36.1-40.7)	33.7%	(29.3-38.4)
Decision Tree	34.4%	(32.7-36.7)	34.4%	(29.9-39.1)
Random Forest	59.1%	(56.6-61.4)	37.2%	(32.6-42.0)

Testing data under both conditions leads to the conclusion that classical modeling methods such as the logistic regression can still be useful for complex modeling problems. Despite this more sophisticated methods that better capture data peculiarities can yield even better results.

Discussion

Results presented clearly show that modeling consumer choice under radical uncertainty is a particular challenge for both economic theory and statistical methods. In the experimental situation we present here consumers have limited idea as to the real dynamics and of the economic system or the probability distribution of expected deliveries. In this case they need to choose either following their intuition or a strategy in order to maximize the amount of omnium, and thus their monetary gain. Different factor tend to affect their choice in a non-linear manner and psychology and emotion can dominate choice. In particular we see a clear S-curve between reported satisfaction from purchases and the difference between promised and delivered quantity across all suppliers and both genders (Figure 1).

As the delta between promise and performance grows wide apart initially agents react by increasing or decreasing their satisfaction ratings. However, beyond a certain threshold level same differences produce a large swing in ratings. This is likely a threshold beyond which emotional factors take precedence in consumer choice. Our first major conclusion is therefore that psychological and emotional factors can dominate choice under radical uncertainty and data for them need to be included in the modeling exercise and accounted for by the method used. The importance of emotional influence in this situation is also confirmed in alternative analysis of this data (Mengov, 2014).
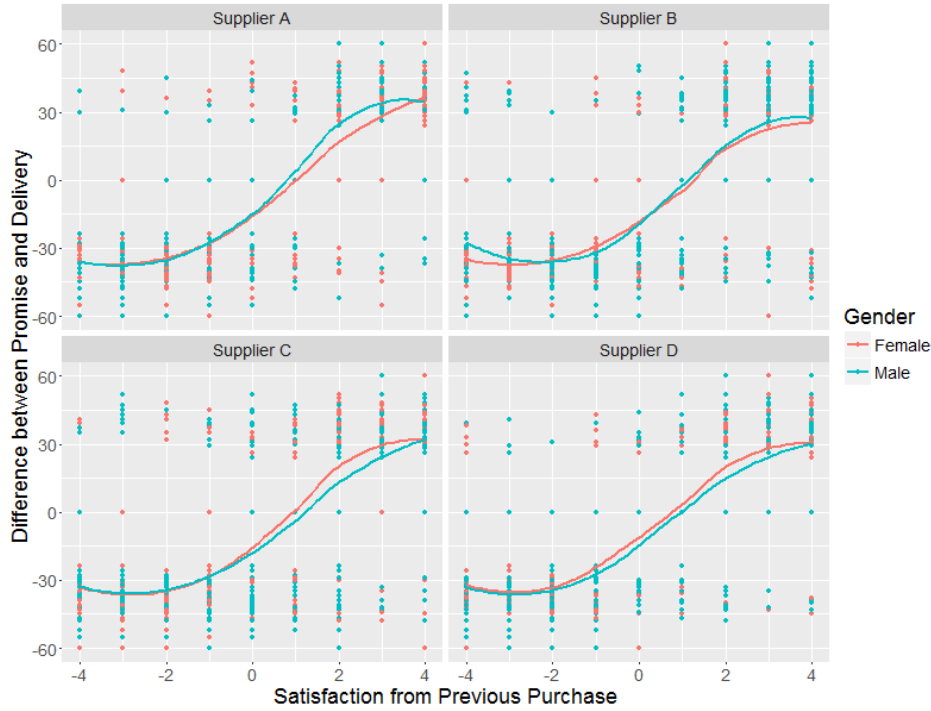
Figure 1: Non-linearity in decisions under radical uncertainty

Additional insight into consumer decision-making can be gleaned by inspecting the variable importance of the best-performing Random Forest model. For this purpose we define the Gini impurity index. If the probability of a given observation y_i to belong to a class C_k out of k classes is denoted as p_k, then the Gini index G is defined as follows:

$$G = \sum_{i=1}^{k} p_i(1 - p_i).$$ (13)

Its mean decrease with the addition of new classification variables can be tracked thus obtaining a measure of variable importance. Here, variable importance denotes what factors contribute most to classification accuracy in terms of the mean decrease in Gini (node impurity) when they are added. There are both similarities and systematic differences depending on whether participants have recourse to a social network or not (Figure 2). Deliberation times (time1), economic conditions (cycle), gender and cumulative result (cum.res) tend to be useful for decision modeling irrespective of whether participants have access to a social network.

Under the conditions without social network agents are very dependent on their own experiences when choosing – their own consumption history (lag1, lag2, lag3) and the experience they had in those choices (satis1, satis2, satis3) in the previous three rounds have large and robust importance. In contrast, these factors lose prominence in the conditions with a social network. Social effects turn out to be crucial factors, with the average reputation variables exhibiting significant importance. In short, personal experience, and particularly consumption habit, is overshadowed by the wisdom of the crowds.

Social influence also modifies the effect of age on decision-making. When no social network information is available, difference in age and thus experience are important for choice modeling. In the opposite condition, other's experience supplants one's own through network effects, and thus age diminishes in significance. Social effects then have a powerful influence on economic choice in such a way that makes individual prediction more difficult for the whole range of examined approaches. This is evidenced by the decreases in prediction accuracy and the increased range of the 95% confidence interval, pointing to a rise in uncertainty.
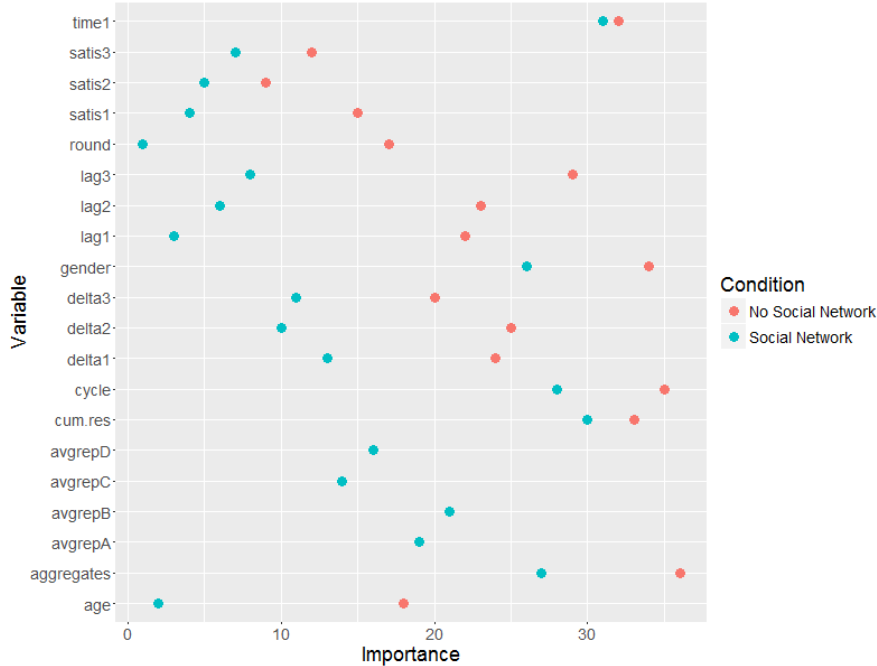

Figure 2: Drivers of Individual Decision Making under Alternative Conditions

A brief inspection of the accuracy tables points towards a methodological conclusion. Such situations of choice under extreme information asymmetry lend themselves best to modeling through advanced machine learning approaches. While traditional methods can still stand their ground, the Random Forest (RF) model fits data better. In addition to that, interpretation is easier under a machine learning algorithm. The logistic regression provides estimates for statistical significance of beta coefficients showing that each of them reached significance well below 1%, which gives only limited information for variable importance. At the same time machine learning algorithms provide a measure of classification improvement which clearly ranks variables. Those benefits only compound at scale – an important consideration in light of exponentially increasing sample sizes.

Conclusion

Gaining a better understanding of how customers chose is an important task with both business and public policy implications. Traditional rational choice economics leaves this topic wanting and empirical models have tended to dominate practical applications. In this paper we study the problem in a complex setting, where participants in a laboratory experiment

need to pick their preferred supplier with limited information about the purchase outcome and under the influence of both cyclical economic factors and randomness in delivery.

Participants are affected by traditional demographics such as age and gender, as well as situational factors like their emotional state and past consumption history. There is one pointed difference in decision-making in the presence of a social network – players tend to substitute their own experience for network information, whereby individual history decreases in importance and reputation, which is essentially an average of other people's satisfaction, gains in prominence. The more complex choice context also makes decision modeling more challenging – something, which is hardly solved through more complicated models. In such situations practitioners should consider collecting more data at a greater level of detail to achieve better classification.

In terms of methods used, such data was best fit by a Random Forest model which is able to account for non-linear relationships between variables and to leverage patterns that are not directly obvious. This modeling approach also provides both an intuitive interpretation and robustly ranks variable importance, which is also of practical significance. This points in the direction that modeling consumption choice will be much helped by both usage of machine learning algorithms and a marked rise in sample sizes.

References

Akinci, S., Kaynak, E., Atilgan, E., & Aksoy, S. (2007). Where does the logistic regression analysis stand in marketing literature? A comparison of the Market Positioning of Prominent Marketing Journals. *European Journal of Marketing*, *41*(5/6), 537-567.

Allais, M. (1953), 'Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'Ecole Americaine', *Econometrica* **21**(4), 503–546.

Bahamonde, A., Díez, J., Quevedo, J. R., Luaces, O., & del Coz, J. J. (2007). How to learn consumer preferences from the analysis of sensory data by means of support vector machines (SVM). *Trends in food science & technology*, *18*(1), 20-28.

Bi, J. (2012). A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *Journal of Sensory Studies*, *27*(2), 87-101.

Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. (1984) *Classification and Regression Trees.* Wadsworth.

Cheung, K. W., Kwok, J. T., Law, M. H., & Tsui, K. C. (2003). Mining customer product ratings for personalized marketing. *Decision Support Systems*, *35*(2), 231-243.

Cortes, C. & Vapnik, V. (1995). Support-vector network. Machine Learning, 20, 1-25.

Ellsberg, D. (1961), 'Risk, Ambiguity, and the Savage Axioms', *Quarterly Journal of Economics* **75**(4), 643–669.

Gan, Christopher, Visit Limsombunchai, Mike Clemes, and Amy Weng. "Consumer Choice Prediction: Artificial Neural Networks versus Logistic Models." *Journal of Social Sciences* 1, no. 4 (2005): 211-219.

Hansen, T. (2005). Consumer adoption of online grocery buying: a discriminant analysis. *International Journal of Retail & Distribution Management*, *33*(2), 101-121.

Hastie, T., Tibshirani, R., & Friedman, J. (2011). *The Elements of Statistical Learning*. NY: Springer.

Hu, M. Y., Shanker, M., & Hung, M. S. (1999). Estimation of posterior probabilities of consumer situational choices with neural network classifiers.*International Journal of Research in Marketing*, *16*(4), 307-317.

Hu, M. Y., Shanker, M., Zhang, G. P., & Hung, M. S. (2008). Modeling consumer situational choice of long distance communication with neural networks. *Decision Support Systems*, *44*(4), 899-908.

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, *39*(1), 1414-1425.

Hyman, M. R., & Yang, Z. (2001). International marketing serials: a retrospective. *International Marketing Review*, *18*(6), 667-718.

Kahneman, D. & Tversky, A. (1972), 'Subjective probability: A judgment of representativeness', *Cognitive Psychology* **3**(3), 430–454.

Kahneman, D. & Tversky, A. (1979), 'Prospect theory: An analysis of decisions under risk', *Econometrica* **47**(2), 263–291.

Kim, J. W., Lee, B. H., Shaw, M. J., Chang, H. L., & Nelson, M. (2001). Application of decision-tree induction techniques to personalized advertisements on internet storefronts. *International Journal of Electronic Commerce*, *5*(3), 45-62.

Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, *40*(13), 5125-5131.

Lee, T. S., Chiu, C. C., Chou, Y. C., & Lu, C. J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, *50*(4), 1113-1130.

Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In *Machine learning: ECML-98* (pp. 4-15). Springer Berlin Heidelberg.

Li, S. T., Shiue, W., & Huang, M. H. (2006). The evaluation of consumer loans using support vector machines. *Expert Systems with Applications*,*30*(4), 772-782.

Mengov, G. (2014). Person-by-person prediction of intuitive economic choice. *Neural Networks*, 60, 232-245.

Neumann, J. v. & Morgenstern, O. (1944), *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ.

Razi, M. A., & Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, *29*(1), 65-74.

Ripley, B. D. (1996) *Pattern Recognition and Neural Networks*. Cambridge University Press.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*(1), 39-60.

Tregear, A., & Ness, M. (2005). Discriminant analysis of consumer interest in buying locally produced foods. *Journal of Marketing management*, *21*(1-2), 19-35.

Xu, K. S., Wang, W., Ren, J., Xu, J. S., Liu, L., & Liao, S. (2012). Classifying consumer comparison opinions to uncover product strengths and weaknesses. *Organizational Efficiency through Intelligent Information Technologies*, 1.

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches.*Expert Systems with Applications*, *36*(3), 6527-6535.

Zhao, X., Shi, Y., Lee, J., Kim, H. K., & Lee, H. (2014). Customer Churn Prediction Based on Feature Clustering and Nonparallel Support Vector Machine. *International Journal of Information Technology & Decision Making*,*13*(05), 1013-1027.