Predicting Consumer Preference From Reviews of Professional Tasting Panels on the Gastrograph Sensory System

June 6th, 2017

Deval Shah, Ryan Ahn, and Jason Cohen
Analytical Flavor Systems, Inc.

Abstract

Analytical Flavor Systems has developed a methodology that allows our artificial intelligence platform to infer the distribution of consumer preference from the reviews of professional panelists on the Gastrograph System.

Sensory panels are a poor representation of general market preference as they do not contain a stratified sampling of the general or target population. Thus, standard statistical methods cannot be employed in order to understand the perception and preferences of the target consumer groups as these statistics do not generalize. This leads companies to developing products that are not competitive with existing products on the market or even products that are outright dis-preferred by the target consumer base.

1 Introduction

The standard sensory practices employed at food and beverage producers aim to ascertain whether a new product will succeed in the general market or with a target consumer cohort, to establish whether the company can improve their existing products, and to determine whether every batch of product is consistent enough to send out. Unfortunately, it is difficult to predict whether they will enjoy the flavor profile of a new product, if that new flavor profile is competitive with the other products on the market, or whether the target population will taste an off-flavor in a deviant batch. This is because sensory panels do not contain a robust and accurate representation of the general population at large. This paper shows a more accurate and predictive methodology to determine consumer preference of any population from reviews of professional panelists.

2 Experimental

2.1 Methods and Data Collection

To project the preferences of professional panelists onto the preferences of the general population, reviews on the Gastrograph Sensory System were transformed to a target experience level and sampled in accordance with the tasting experience level distribution of the general population. The techniques LFDA (Local Fisher Discriminant Analysis) and PAM (Partitioning around Medoids) were used to maximize between-product similarity and minimize within-product similarity in flavor profile. The random forest method was then utilized to predict the distribution of perceived quality scores the general population would assign given any set of reviews, with built-in considerations for the style of beer and the tasting experience of the panelists.

The Gastrograph Sensory System is a methodology for the robust collection and aggregation of sensory data on Gastrograph Review for Android and iOS. The Gastrograph system asks reviewers to rate the intensity of 24 broad spectrum flavor attributes and somatosensations on a scale of 0 5 (0 being not present at all, 1 being at the edge of perception, and 5 being the highest possible intensity in a product), supplemented by the specific reference flavors the reviewers taste in that product categorized within a single broad-spectrum flavor attribute. At the end of each review, the reviewer is asked to assign a perceived quality score between 1 and 7 to rate their assessment of the products quality. Each panelist has an account on the Gastrograph Sensory System allowing the tracking and controlling of their respective biases[1] and flavor sensitivies[2]. The users are assigned a per-review experience score indicating how likely it is that their review is an accurate representation of the product in question.

2.2 Experience Score

As cognizant tasting becomes more frequent, perception changes, and preferences evolve. The ability of the panelists to identify subtleties and nuances in flavor that one would not have identified in the past is quantifiable and mappable. Experience score is a metric used to determine the panelist's proficiency at that review in identifying the subtleties of the flavor profile in a product across multiple somatosensory attributes [3].

All panelists, trained or untrained, at any experience level, are themselves also consumers. As a panelist gains experience, their perception changes with the increased ability to identify subtlety and nuance in a products flavor profile distancing their perception from that of a consumer panel. This is a problem that must be solved in order to predict consumer preference from panelist reviews.

We solve this by transforming reviews using locality and covariance preserving projections from any given experience level to the target experience level or distribution of target experience levels. This is a viable way to infer hedonic preferences of sensory attributes as untrained panelists at lower experience levels found in the general or target population being modeled are equivalent to consumers. We can further show that the Perceived Quality scores of panelists with low experience are indistinguishable from hedonic enjoyment scores to the average consumer, the quality of a product is how much they prefer that product.

Thus, the equivalent low experience consumer data can be used as a training set to learn the projection for predicting perception of a product from higher experience reviewers to consumers at any experience level. Once this is done, data can be generated for any distribution of experience scores from professional panelist data.

2.3 Style of Product

To predict how the product will fare in any given market, reviews are clustered into styles based on their flavor profiles. This is done by projecting the dataset of beer reviews into new dimensions using Local Fisher Discriminant Analysis, which minimizes within-product scatter, and maximizes between-product scatter, while preserving local relationships between reviews. The PAM (Partitioning Around Medoids) algorithm is then used to cluster similar products together[6]. The sets of products in the same cluster are considered to be part of the same style.

2.4 Predicting Market Preference

To model reviews of consumers in the general population, Gastrogaph AI transforms the reviews of the product of interest using locality and covariance preserving projections to match a stratified sampling proportional to the known distribution of experience scores of the population of interest. The experience score metric accounts for previous tasting experience for new users on the Gastrograph Sensory System, so an accurate sampling is obtained from the experience scores the users started with [3]. This results in a more accurate portrayal of the tasting experience of the general population because most consumers are of a lower experience score than an individual on a companys sensory panel.

Random forest is a machine learning algorithm that uses hundreds of decision trees, each with a subset both of the variables and of the observations in the input data, to both predict the output perceived quality and learn the variables of most importance. Decision trees are a set of rules used to classify the data into categories. In this case, the categories are the different possible perceived quality scores on a scale of 1 to 7, 7 being the highest. The variables chosen to categorize each observation are the ones that maximize information gain when used as a splitter in the decision trees. In this case, information gain is determined using Shannon entropy, which is calculated by the formula

$$H(X) = -\sum_{i=1}^{N} p_i log_2(p_i)$$

where i is a possible category (in this case, the perceived quality), and p_i is the probability of that category appearing in the data. For instance, if the dataset is split equally between perceived quality scores of 1 and 2, 50% of the observations are in one subset and 50% of the observations are in the other. The entropy is then $-((.5 * log_2(.5)) + (.5 * log_2(.5)))$, which is equal to 1 (the maximum possible amount of disorder). If t denotes the system before a split, and $t+1$ denotes the system after a split, information gain is $H_t(X) - H_{t+1}(X)$. In other words, information gain is maximized by maximizing the decrease in entropy. In this manner, the splitting variables are chosen.

The random forest algorithm is run with the inputs being the flavor profiles, the experience score, and the style classification of each review of beer, and the output being a probability distribution of possible perceived quality scores the review might be classified under. The random forest model is trained on the flavor profiles of every review, to predict the output the probabilities of perceived quality given to the review. A log loss function is used as the error metric.

The log loss function is

$$-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} log(p_{ij})$$

where N is the number of observations, M is the number of possible values for the output, y_{ij} is a binary indicator of whether or not j is the correct output for the input i, and p_{ij} is the probability of the model assigning the label j to the input i. The slope of the function is steeper at the extremes of probability outputs, so if the model ever assigns a low probability value to a plausible perceived quality score, the error goes up significantly.

This loss function has the effect of heavily penalizing a near zero probability assignment of label j to any input i, which biases the output probabilities away from complete certainty. Therefore, the model will predict the distribution of perceived quality scores for a review or set of reviews, with a higher variance of potential perceived quality scores to account for the inherent variability of preference. However, as the experience score of the review increases, the variability of preference decreases, and the perceived quality scores converge considerably.

3 Results and Discussion

The output of this model is a distribution of perceived quality score probabilities for any input review(s). These probabilities are calculated from the proportion of decision trees that voted for this perceived quality score. These probabilities are interpreted as the predicted market preference of the input reviews. This is a valid interpretation because of the correct distribution of experience scores in the input dataset, and the understanding that the decision trees model various segments of the population that do not care about each flavor attribute equally. Because the style classification of each product is included in the model, the model learns to discriminate between product or styles of products in a more robust way: for example the combinations of flavors that make a stout achieve a high perceived quality score are different than the ones that make an IPA achieve a high perceived quality score.

The model achieves a mean absolute error of 0.98 perceived quality units on our set of test data, which is lower than the standard deviation of 1.3 perceived quality units in our dataset. This amount of error is expected to remain due to the inherent variability of perceived quality, especially when taking into account less experienced reviewers. The result is a robust model of the distribution of perceived quality scores that the general population would taste, given any number of input reviews.

Calculating the predicted percentage of the population that would assign a product each possible perceived quality score is done by taking the max value of the integral of the probability density function of the predicted market preference of one set of reviews minus the probability density function of the predicted market preference of the other set of reviews, from the perceived quality of every possible point to 7. In mathematical terms,

$$P = \int_{argmax_{x \in (1,7)}(\int_x^7 f_a(x) - f_b(x)dx)}^{7} f_a(x) - f_b(x)dx$$

where P is the percent of people that prefer product a, $f_a(x)$ is the probability density function of the predicted market performance of the review set of product a, and $f_b(x)$ is the probability density function of the predicted market preference of the review set of product b. With this information, brewers are now able to determine what percentage of the population will prefer the shipment of an either positively or negatively deviant batch. This tool further allows brewers to guide new product development and existing product optimization in a data-driven way: understanding, quantitatively, how a change in the flavor profile of a product affects the market preference, and therefore performance, of that product.

4 Applications

Predicting the percentage of the population that would prefer an existing product, a product under development, or even a specific batch of a product (versus another batch or the overall product) is now immediately quantifiable and can help companies predict market performance, guide innovation strategy, and ensure that quality products are being produced. Other applications of this come from subsetting the population by gender and quantifying that difference in predicted market preference for any product, and formulating new products with a target flavor profile that results in successful market performance. This can be done for any class of product in the food and beverage space, given sufficient data.

5 Conclusion

A robust methodology for measuring and predicting consumer preference from professional panels is now proven possible with the Gastrograph System. The predictive capabilities of Gastrograph AI is already helping companies develop new products, optimize existing brands, target their highest value consumer cohorts, and improve the quality at their production facilities by monitoring batch-to-batch deviations. Historically, consumer preference is notoriously difficult to predict due to factors including the rise and fall of many market trends and changing consumer preferences. This study introduces a novel way to understand and predict market preference using machine learning from the reviews of panelists.

Literature Cited

[1] Gupta, Rishab and Jason Cohen. *"Determining Flavor-Attribute Sensitivity Of Panelists"*. Gastronexus. N.p., 2016. Web. 27 Jan. 2017.

[2] Leon, Arturo and Jason Cohen.*"Quantifying Reviewer Trust: Identifying Reviewers As Trustworthy Through Knn"*. GastroNexus. N.p., 2015. Web. 27 Jan. 2017.

[3] Leon, Arturo and Jason Cohen. *"Classifying Reviewers By Experience: Modeling The Evolution Of Reviewer Experience"*. GastroNexus. N.p., 2015. Web. 27 Jan. 2017.

[4] *"Making Sense Of Logarithmic Loss — Exegetic Analytics"*. Exegetic Analytics. N. p., 2013. Web. 13 Jan. 2017

[5] Sugiyama, Masashi. *"Local Fisher Discriminant Analysis For Supervised Dimensionality Reduction"*. n. pag. Print.

[6] Yuan, Tang. *"Local Fisher Discriminant Analysis On Beer Style Clustering"*. GastroNexus. N.p., 2014. Web. 27 Jan. 2017.

6 Appendix

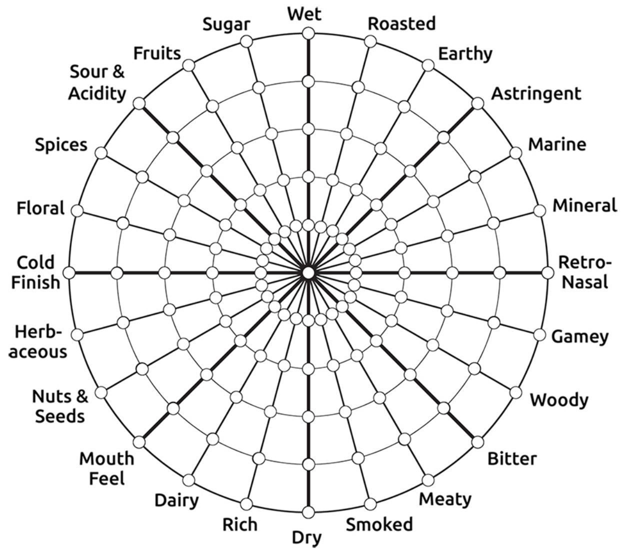

Figure 1: The data collection interface of the Gastrograph Sensory System

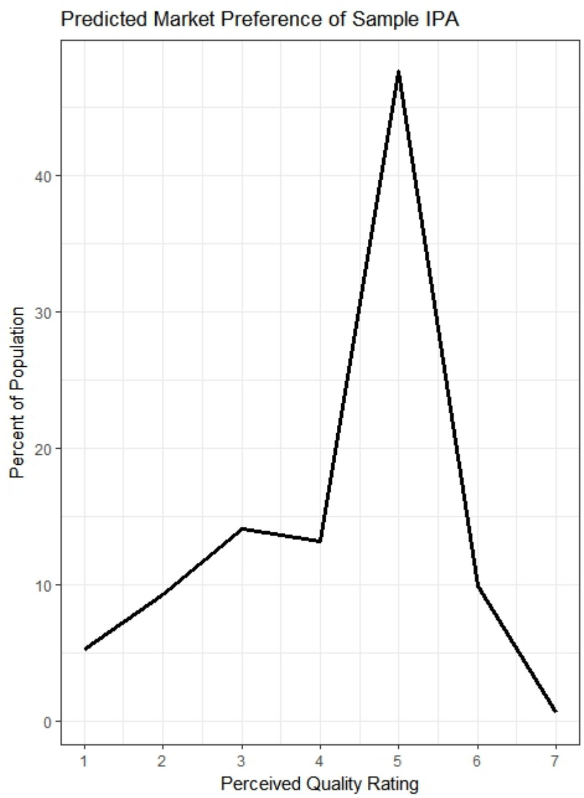

Figure 2: The predicted preference distribution of the general beer drinking market for a sample IPA