Assessing the Value of eBay Listing Features

Leon Lin - (leonlin@stanford.edu)
Vasant Ramachandran – (vasantr@stanford.edu)

December 10, 2010

Abstract

We used machine learning to access the value of additional features sellers can use to highlight their eBay item listings. The set of features a user can add to their listing includes subtitles, extra photos, a pop up photo view of an item, a "listing designer", and a bold listing of their item in search results. As eBay fixes the costs for these features, we implemented a price-prediction scheme to determine which listing features add most value to an item and identified the relative importance of specific features in price determination.

Motivation

In 2005, online marketplaces such as eBay, Yahoo! Auctions, and Amazon Marketplace accounted for 25% of all online e-commerce [1]. With over $32 billion of product sales and over 430,000 registered users considering revenue made from online auctions their primary source of income in 2003 [3], eBay also serves as a significant source of revenue for a vast portion of users. While there is a large component of uncertainty caused by human behavior in predicting the final prices of eBay listings [1], numerous studies have been performed suggesting that there is a high correlation between certain features of an eBay listing and the final buying price of the listing. [4] demonstrates a machine learning algorithm incorporating decision trees which was able to predict the final buying price within 20% up to 89% of the time on certain item keywords. As a result, knowing which features of a product listing to include can be a highly profitable asset for eBay sellers, especially the over 430,000 sellers who in 2004 cited eBay as their primary source of income [3]. The goal of this project is to access the validity of this notion, thereby drawing conclusions regarding what is the optimal feature set to purchase when listing an item, as well as what is the optimal pricing scheme for listing features.A substantial amount of work has been done on analyzing auctions using data mining and statistical techniques. Perhaps most related to our current study is [4] which makes price predictions based upon the number of pictures, feedback rating, and description of the item. To our knowledge, there has been no work specifically focused on the influence of item listing features which sellers can add for a fee on price prediction. Indeed, such prior research seems highly unlikely given that eBay did not define their current set of fee based listing features until October of 2009 [9].

Feature Set and Data

The fixed price addable features are described by eBay as a *Subtitle($0.50)* to capture the interest of buyers when they view search results by displaying more information below the title, *Gallery Plus($0.35)* which displays 400-pixel pictures from your listing in search results when buyers scroll over a Gallery Picture, *Pictures($0.15 per photo--first photo is free)* which allow potential buyers to see more images of the item being sold, a *Listing Designer($0.10)* to enhance item descriptions with an attractive theme that complements the type of item being sold, and *Bold($2.00)* to make the listing title bold in search results.

Thus, in assessing the influence of each of these features on the final selling price of an auction, we were able to draw conclusions regarding the optimal set of features to purchase for an item listing. We used our own web crawler written in Java to extract the features of particular completed eBay listings and manually checked each item listing to insure the accuracy of our training data. We focused specifically on listings returned by a search for "garmin nuvi205w" and "garmin nuvi 260wt," under the "GPS Systems" category on eBay with the specified condition, "new." These items were selected on the basis that they were very popular items and had high keyword specificity. We represented a feature vector as a set of four Boolean values (to indicate presence or absence of a feature) and an integer to represent the number of photos.

Research Methodology

In order to achieve our goals of predicting the end-price of a listing of the item based upon an observation of the features associated with the listing, as well as understanding the relative importance of each feature, we defined our predictions as a choice of one of ten $5 price intervals (buckets) between set minimum and maximum prices. Since most of our features were categorical, this multinomial outcome generation allowed us to compare the relative merits of three separate approaches to price determination: a decision (CART) tree approach, softmax regression, and a naïve-Bayes classifier system to sort observations into buckets.

Classification and Regression Trees

To compare the results of traditional machine learning algorithms to a decision tree approach to the problem, we trained a CART (classification and regression tree) on the input data. For our categorical/numeric variable input set, the split criterion looks like $x_j \in V$, with $V \subset W_j$. In our model, W_j refers to the collection of all possible categories of variable x_j, which for the gallery, subtitle, designer listing, and bold options is just 0—indicating absence—and 1—indicating presence. The "number of photos" variable takes on multiple integer values. The terminal nodes (leaves) contain a $\hat{y}$ value, an estimate for the price, in the leaf. In practice this value is taken to be the average of all observed y values whose decision path ends in the leaf [3]. Once a prediction is made (following a particular decision path to a leaf node), we place the leaf estimate in its price bucket.

Because a decision tree model typically overfits the data, we introduced a pruning phase which used 10-fold cross validation on our training set to optimize our tree to a minimal-cost CART. The cost of the tree is the sum over all terminal nodes of the estimated probability of a node times the cost of a node, while cost of the node is the average squared error over observations at the node[3]. Specifically, the cost represents the inaccuracy of the absolute price prediction when compared to the actual price, rather than the discrepancy between the price bucket prediction and the actual bucket. In our case, we averaged the cost of each of the ten CARTs built during the cross validation process, and sought to minimize this cost function iteratively in a MATLAB routine by successively pruning nodes and calculating subtree costs to achieve the optimal pruning sequence. An estimate for the best level of pruning is defined as the smallest tree that is within one standard error of the minimum-cost subtree[3]. While regression trees can be represented as an amalgamation of simple if-then rules, a stronger relative importance metric for each feature was desirable in order to determine a weight for each feature. For a single CART model, the following formula measures the importance of variable x_j[4]:

$$I_j^{\,2}(B) = \sum_{n=1}^{K-1} (i_n)^2 \, \chi(\mathrm{v_n} = \mathrm{x_j})$$

The summation covers the non-terminal nodes in tree B, which has K leaves. $\chi()$ denotes the indicator function, while ν_n is the split variable of node n (the feature on which the split is made). In order to measure the relative importance of the j^{th} feature, the improvement in average squared error as a result of all splits on the j^{th} feature is calculated.

The factor $(i_n)^2$ measures the improvement in squared error as a result of the split in node n.

$$(i_n)^2 = (y_l - y_r)^2 \, \frac{w_l \, w_r}{w_l + w_r}$$

Here w_l and w_r are the probabilities that the decision path turns to the left or right child node of node n, and y_l and y_r are the average price predictions of paths going through the left and right child nodes[3]. Once the CART is constructed and pruned to optimally, the relative importance metric $I_j^{\,2}$ for price prediction can be calculated for each feature. To achieve a large $I_j^{\,2}$ the feature must be used in several even splits with a significant difference in the average price prediction down all possible left and right paths[3]. Finally, the $I_j^{\,2}$are normalized in our model to an average feature weight of 1.

$$w_j = \frac{5}{\sum_{j=1}^{5} I_j^{\,2}} I_j^{\,2}$$

Multinomial Logistic Regression

In place of a linear or logistic regression model for price prediction, the CART is able to determine with greater flexibility which features to use in order to predict the target variable (price), even if the relationship between price and the feature set is nonlinear. Also, the CART is capable of handling categorical variables and missing values if an untransformed data set is used to train or test [4]. The drawback of the CART model is stability: our data set size is limited, and the model responds adversely to slight changes to the data set (removing duplicate listings, combining listing sets for very similar products, etc)—which necessitates consideration of alternative models. Absolute price prediction with a linear regression model performs poorly on a numeric transformation of the observation feature set, as the set consists of Boolean and discrete integer variables. However, the price "bucket" prediction model works well with multinomial classification and regression--treating each bucket as an outcome [10]. To perform multinomial classification of the observations into price buckets, a multinomial logistic regression model was trained on the data. The weights B_j for each feature are determined on the training set to maximize the log-likelihood of the training observations under the softmax regression model, and then each test observation is classified to the price bucket with the largest probability[10]. This "maximum entropy model" does not assume feature independence.

$$\Pr(y_i = j) = \frac{e^{X_i B_j}}{1 + \sum_{j=1}^{J} e^{X_i B_j}}$$

Multiple Naïve Bayes Classifiers

We created multiple binary classifiers assuming feature independence, with each classifier learning whether the end-price of the auction would be greater than the minimum of a specific price interval or not. The predicted interval is then selected by combining the results from all the classifiers. Inconsistencies between classifiers that should predict the same outcome on test data are resolved in favor of the lowest possible interval. This technique was motivated by the scarcity of training examples for any one specific item in the online auction, as binary classification would be more robust under a sparse training set than a more complicated classification.

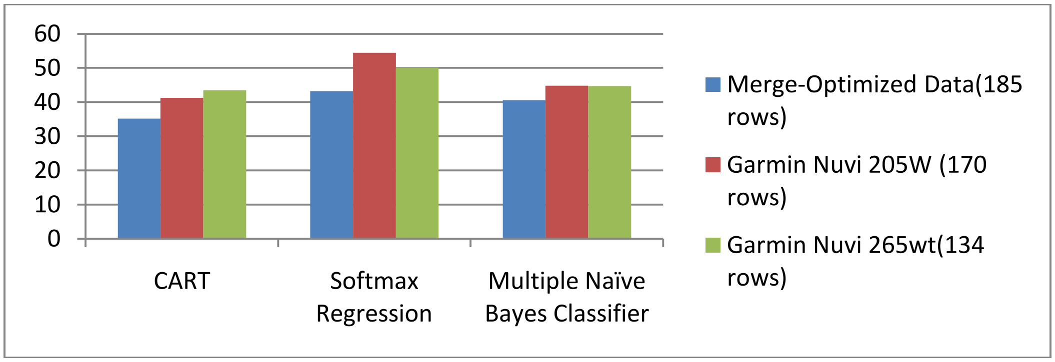
Optimizations:

Many large eBay sellers will tend to sell a particular item in bulk, thereby listing the same item multiple times with the same feature set. Thus, to prevent highly replicated listings from skewing our model parameters, we performed duplicate elimination, where all items with the same title, features, and seller were condensed into one listing. The selling price of the condensed listing was the average price of the replicas.

To combat the problem of limited training data for each model, we decided to create models which combined the training data from both our queries. Performing such a merger, however, presented a trade-off: while we would have more data with which to train our model, we introduce a new component of variability between each of the individual listings. In practice, the variability turned out to be minimal, as the items are very similar. We merged the two data sets by first taking the average final selling prices of the results retrieved from the two keywords, respectively, with duplicates removed, and found that the "garmin nuvi 265wt" sold at an average price of $120.93 while the "garmin nuvi 205w" sold at the average price of $86.81. Upon transforming the 205w data set in this fashion, we found that the new distribution had a variance of 134 upon removal of 3/135 of the outlying data points. The "garmin nuvi 265wt" data set had an equal mean to the transformed data set and a variance of 124. Therefore, assuming that features influence the selling price of twin items in similar ways, we concluded that a merger between the two data sets was now possible.

Results and Analysis

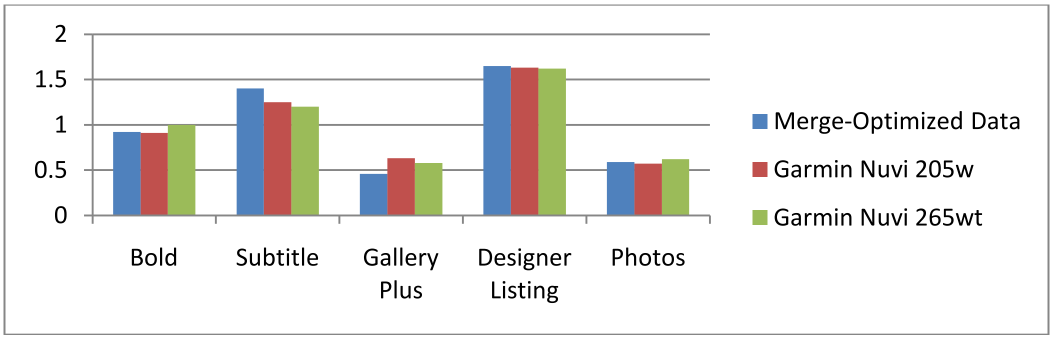
10-fold Cross Validation Error in Price Bucket Prediction(%)



Our intuition about the use of duplicate elimination to reduce cross-validation error was proven correct, as was our choice to merge data sets. Softmax regression exhibits significantly poorer results than the other models, partially due to the use of discrete (integer) features, which do not translate as well as expected to the continuous probability prediction that the multinomial logit requires. Additionally, if the multinomial logit is used to model choices, it assumes independent irrelevance of alternatives (IIA) which is not always desirable [10]. This assumption states that the odds do not depend on other alternatives that are not relevant, but this is not true with market psychology, as buyers typically violate rational choice theory (i.e. the choice between a bold, subtitled car, a subtitled car, and a subtitled bus)[10]. Also, the collinearity assumption (that the attributes are not linearly related) in a regression model is hard to apply to the discrete case, where "presence" implies identical values and unintended correlation [10]. Due to the use of the optimal pruning scheme, the

CART, which was the predictably best performer, does not over-fit as much expected and appears relatively stable under the “merge” of the two training sets. The Naïve Bayes classifier exhibited a training error of only 15-17 % on all three data sets, but a significantly higher cross validation error was observed due to potential overfitting. Due to the high variance of the data, there were only nine to ten occurrences (in cross validation) of inconsistencies in classification (where a classifier with a lower “threshold” would predict a 0 while a “higher” threshold classifier would predict 1), suggesting clear stratification of the pricing.

Relative Importance of Features(CART w_j)



The relative importance of each feature is also presented above. Despite eBay’s pricing scheme, our findings suggest that the designer listing feature was the most important feature for determining the final selling price of an item, followed closely by the subtitle feature while additional photos and Gallery Plus proved to be the least important feature. The significance of designer listings can be explained by the fact that the designer listing feature offer the largest visual impact of all the features, giving the seller complete control over the HTML displayed in the item listing. This might prove particularly significant to the keywords being studied in this paper, since potential buyers of GPS systems often care most about features and specifications, which can be most effectively presented to the user in tables and charts provided by the designer listing feature. At the same time, potential buyers of new GPS systems are likely to be less swayed by the presence of additional photos, since the user is not likely to gain much information from seeing additional photos of the product. This holds especially true for our data set, where we intentionally tried to minimize the item variability amongst different listings. Therefore, for our data set in particular, listings are unlikely the capture the attention of users by displaying additional photos of the item to users because all photos illustrate the same item. What the seller is truly able to customize is, instead, the actual presentation of the item and its description and specifications, which can be significantly enhanced through the listing designer.

Human vs. Machine Learning

A 10-fold cross validation on the CART cost (which calculates squared error in absolute price prediction across the tree) showed us that the error is comparable in value (when normalized to a percentage) to the bucket prediction cross validation error shown above. Thus, the bucketing did not introduce significant error to our CART model, which was contrary to expectation [4].

Another hypothesis we wanted to test was the correlation between seller feedback score and the total amount that a seller spent on listing features when using eBay’s pricing scheme vs. using a pricing scheme that was deemed optimal based upon the weights outputted by our model. For this pricing scheme, we set the feature with the highest weight at maximum price and scaled the other feature prices accordingly. Over time, we hypothesized users with higher feedback scores will tend to purchase features which carry the most weight, and thus will have the highest probability of increasing the final selling price of the listed item. We found that there was higher correlation between the log of seller feedback scores to money spent on listing features when pricing by our optimal weight scheme vs. the log of selling feedback scores to money spent on features when using eBay’s pricing scheme. While the correlation between feedback score and price spent on listing features was relatively small in both cases, it is nonetheless interesting to note that experienced sellers tend to select listing features which correspond more closely to the strategy suggested by our weighting system, rather than the system proposed by eBay’s current pricing scheme. Therefore, sellers on eBay are “learning” the same model as our machine learning algorithms are.

Conclusions and Future Work

The largest limitation of our project was the scope of our training set. While we would have liked to utilize a much larger and diverse training set, our abilities to procure a larger training set were curtailed by our inability to automatically retrieve feature information for each item. eBay blocks any scraping of completed listings, and so our feature vectors had to be manually created. This only allowed us to train on a data set of 304 listings in total due to time constraints with fetching the data set. Additionally, we were only able to train on models on items from a specific category, which is quite a limiting constraint given that feature weights would likely vary heavily depending upon the category being investigated. For instance, the presence of photos would probably be much more important when purchasing rare paintings vs. purchasing a new electronic item, since electronic items are the same from listing to listing when searched under the same keyword. Nevertheless, we were able to establish a sense of relative importance for each feature, a relatively accurate price-predictor, and an optimal pricing scheme for the feature set.

References

1. Ghani, Rayid, and Hillery Simmons. "Predicting the End-Price of Online Auctions." *Accenture Technology Labs* (). Web. 12 Nov. 2010.
2. "Number of active eBay Sellers ." *Google Answers*. eBay, 20 Jan. 2010. Web. 24 Nov. 2005. <http://answers.google.com/answers/threadview?id=596972>.
3. Heijst, Dennis v., Rob Potharst, and Michiel v. Wezel. "A support system for predicting eBay end prices." *Decision Support Systems* 44 (2007): 970-82. Web. 12 Nov. 2010
4. L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, 2nd Edition, Chapman & Hall, NY, 1996.
5. D. Bryan, D. Lucking-Reiley, N. Prasad, D. Reeves, Pennies from eBay: The Determinants of Price in Online Auctions. Working Papers 0003, Department of Economics, Vanderbilt University, November 1999 (May 2006 Version).
6. A.E. Roth, A. Ockenfels, Last-minute bidding and the rules for ending second-price auctions: evidence from eBay and Amazon auctions on the internet, American Economic Review 92 (4) (2002) 1093–1103 September.
7. Shanshan Wang, Wolfgang Jank, Galit Shmueli. Explaining and forecasting online auction prices and their dynamics using functional data analysis. Journal of Business and Economic Statistics, in press.
8. "Seller update: September 2009." *eBay*. eBay, n.d. Web. 12 Nov. 2010. <http://pages.ebay.com/sell/July2009Update/Details/index.html#2-4>.
9. "Promoting your item with listing upgrades." *eBay*. N.p., n.d. Web. 9 Dec. 2010. <http://pages.ebay.com/help/sell/promoting_ov.html>.
10. Hilbe, Joseph M. (2009). *Logistic Regression Models.* Chapman & Hall/CRC Press.