CSCI-5622: When They Buzz

Alex Gendreau   Garrett Lewellen   Tyler Behm

May 7$^{\text{th}}$, 2015

Problem

Contestants in a quiz bowl may answer a question as it is being read by buzzing in. We wish to predict when this will happen so that an artificial agent will buzz in just before other contestants. We treat this as two separate problems 1) at what point in the question will the user buzz in, and 2) will user answer correctly. Correctly answered positions will be reported with a positive sign, and incorrectly negatively. We measure (1) by root mean square error (RMSE) and (2) by sign accuracy (ACC).

Original Approach

We originally proposed a user-dependent probabilistic model based on independence assumptions to pick the position with the highest probability given question features. Using logistic regression we found the approach slow and inaccurate (RMSE 106). We pivoted in favor of two simpler parallel lines of work: the Expected Value approach and Right-Wrong approach.
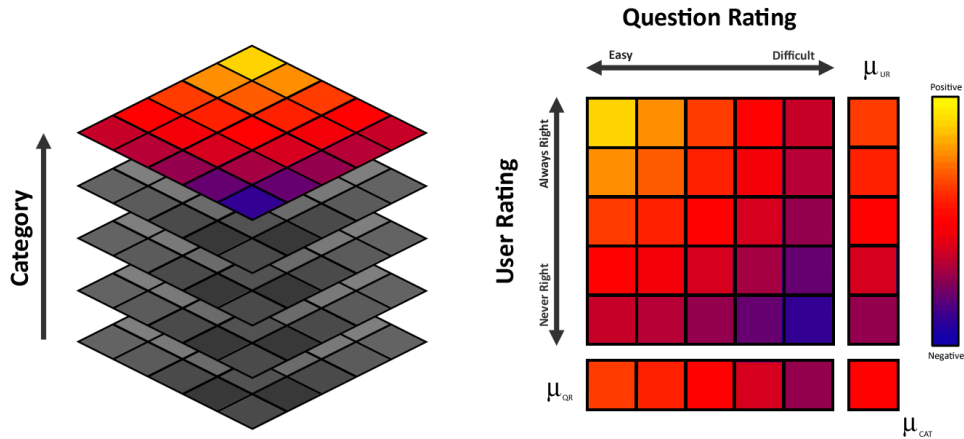
Expected Value Approach

Figure 1: Visual representation of the approach. If user rating or question rating are unknown, then expected positions μ_{QR} and μ_{UR} are used, or μ_{CAT} if neither is known.

Garrett's simplest approach reported the expected position based on category, user rating, and question rating as depicted in Fig. (1). Users' abilities are measured by the portion of questions they answer correctly denoted by an integer value in $[-7, 7]$. Questions' difficulties are similar, but measured by

portion of users who answer them correctly denoted by an integer in $[-2, 2]$. Garrett believed these features ultimately influenced the resulting sign and position. Result: RMSE of 87.48 and ACC of 72.50%.
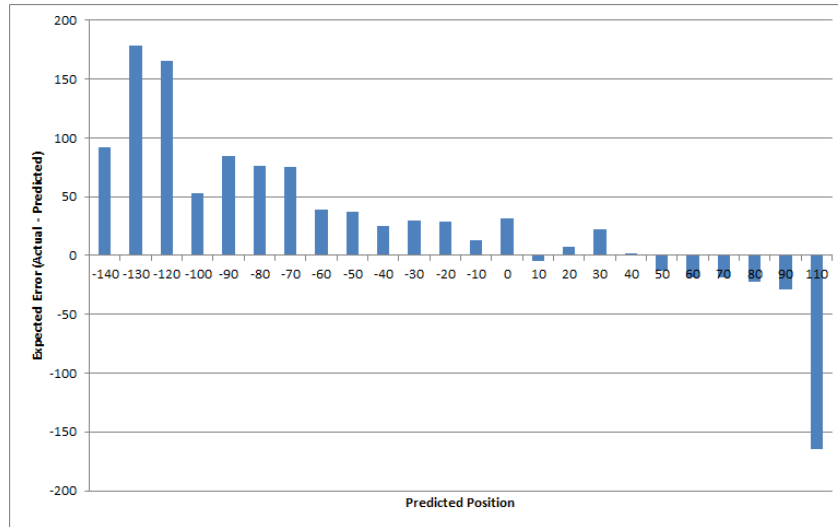

Figure 2: Expected error as a function of predicted value.

To improve RMSE, Garrett looked at the relationship between expected error and predicted position. Applying these error corrections to subsequent predictions produced RMSE of 84.39 and ACC of 73.02%. Repeating the process three more times yielded RMSE of 83.73 and 72.65%. Next, he looked at the most frequently reported position, user rating, and question rating to apply further adjustments based on category to obtain RMSE of 83.25 and sign accuracy of 73.5%.
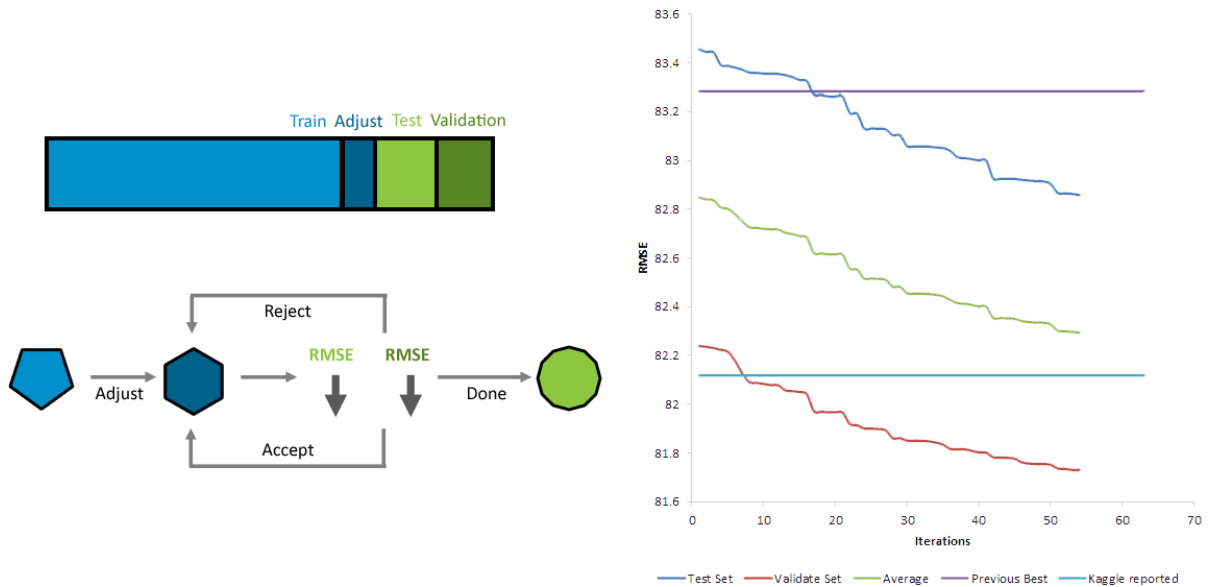

Figure 3: (Left) Visual representation of refinement algorithm. (Right) RMSE as a function of algorithm iteration.

Finally, he automated this adjustment process as depicted in Fig. (3) by splitting the training set into four sets: two to initially fit and adjust the model, and two to verify adjustments reduced RMSE on two separate sets to mimic the Kaggle setup. An adjustment is applied to the model, if RMSE does not improve for both test sets, the adjustment is rejected, otherwise it is accepted, and the process continues until the adjustment set is exhausted. Result: RMSE of 82.11832 as our best Kaggle score until May 2nd.

Right-Wrong Approach

Tyler TODO (Sign, then position & position then sign, features, error analysis)

Final Approach

Alex TODO (Hedge bets, continuous yes/no sign prediction, features, error analysis)

Conclusions

Sunday we obtained our best score of 81.55246 based on the final approach. ...