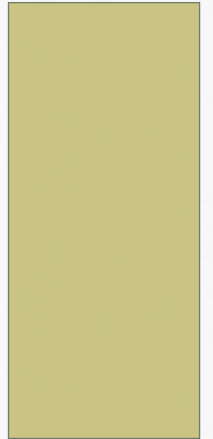


CAPITAL ONE DATA MINING CUP

TEAM MEMBERS:

RUIZHE LUO, MENGYE REN, GEYA XU, YUHAO ZHAO



OBJECT AND MAIN PROBLEM

- **Object:**

Predict conversion rate and break-even bids from given information.

- **Main Problem**

Conversion Rate

- Predict conversion rate
- Decide Approval rate

Revenue

- Product specific average revenue
- Expected product specific revenue given applied

DATA MODIFICATION

- build_set : the data used to train the models with NA removed.
- valid_set: the data used to predict the conversation rate and break-even bids with NA removed.

KEYWORD-PRODUCT RELATION

- Projected the 492 given keywords onto 50-dimension vectors
- Sum all keywords in a bag
- Applied logistic regression for each of the 6 individual product (given at least one of the products has been applied)
- Added 5x more penalty on false negative examples than false positive examples
- Trained for 300 epochs
- 79% accuracy and 77% recall rate

MAIN STRATEGY

Model for conversation rate:

- Learn the relation between keywords bundles and individual products that were applied using logistic regression. (output (before normalization) was given by KEYWORD_TO_PRODUCT_BUILD.csv and KEYWORD_TO_PRODUCT_VALID.csv)
- Build a linear model “*predapp*” to predict the total application by inputs VISITS, ENGN_ID, LANG_ID and DVIC_ID
- Predict application rates from the valid_set, by multiplying the fitted applications to the normalized application distribution from the keyword learning .Assign 0 to rows with zero application or zero visits. We treat the application rates as the conversation rate. Note that the conversation rate is a matrix with 6 columns.

MAIN STRATEGY

Model for Approve Rate:

- Model the average approve rate for each product (a 6 dimension vector)
- Approve Rate for each product is modeled by:
predicted application rate \times Product average approve rate

Model for Revenue:

- Learn the product specific average revenue from the build_set
- Expected product specific revenue given applied is modeled by:
predicted approval rate \times Product average revenue

Then : $\text{MAX_BID} = \text{Revenue} \times (\text{Conversation Rate} * \text{A/R})$