Predicting the Value of Jeopardy Questions

SI 670: Ermioni Carr, Yashaswini Joshi, & Hsin-Yuan Wu

Motivation

Our goal for this project was to predict the value of jeopardy questions based on each questions' text and answer. If successful, such a model could be used to verify or assign values to future questions. Also, we would gain knowledge regarding what aspects of a question make it more "difficult" and if/how these can be represented as model features.

Methodology

Data Collection → Data Cleaning and Analysis → Data preprocessing & Feature Extraction → Data Modeling

Data Source:
- Jeopardy Dataset with 216,930 rows
- Columns include the category, value, question text, answer text, round, show number and air date

Cleaning and Filtering:
- Only 'Jeopardy!' round
- No answers w/ numbers or questions w/ hyperlinks
- Only show #'s after 4000
- Only questions with >= 5 words
- Random subset of 2,000 cases for each value ($200, $1000)

Text Pre-processing:
- Stop word removal
- Tokenization

Feature Extraction :
- TF-IDF
- Cosine Similarity
- Textstat library
- Word Embeddings
- Word Mover's distance
- Custom functions (e.g., longest word)

Baseline Models:
- Dummy (Majority) Classifier

Preprocessing
- MinMaxScaler
- SimpleImputer

Predictive Classifiers:
- Naive Bayes
- Random Forest
- MLP
- SVC
- LSTM

Acknowledgements

We would like to thank the teaching staff and Prof. Grant for their help and support.

Results

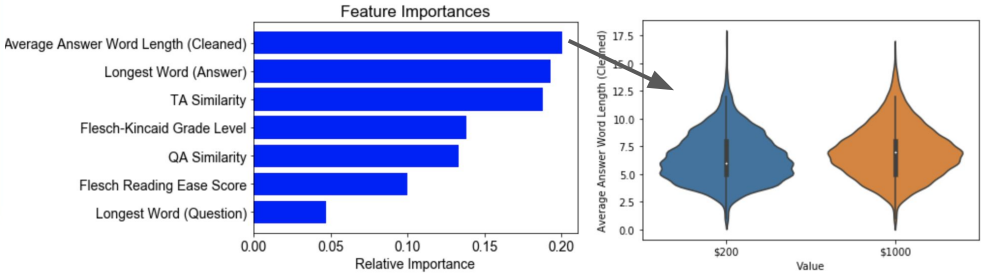
Model Comparison *(Feature set: 'Flesch Reading Ease Score', 'Flesch-Kincaid Grade Level', 'Longest Word (Question)', 'Longest Word (Answer)', 'Average Answer Word Length (Cleaned)', 'QA Similarity', and 'TA Similarity'):*

	SVC	Naive Bayes	MLP	RF
Training Accuracy	57.1%	56.6%	58.1%	**58.6%**
Testing Accuracy	56.4%	55%	57%	**58%**

Best Model: Random Forest *(n_estimators=100, max_features=2, max_depth=2)*

Baseline Accuracy (Dummy Classifier) = **50%**
Training Accuracy = **58.6%**
Testing Accuracy = **58%**

	precision	recall	f1-score
$200	0.58	0.61	0.59
$1000	0.59	0.55	0.57



Conclusion: We saw that all of our models had poor performance on this task. We hypothesize that this is related to the nature of the task: assigning monetary values to questions is subjective and dependent on domain knowledge.