Detection of Dialogue Act

Data Science Institute
COLUMBIA UNIVERSITY

Huanyu Jiang(hj2593), Binghong Yu(by2325), Huaizhi Ge(hg2590), Keyi Guo(kg2955), Anbang Wang(aw3396)

Industry Mentors: Ivan Wong, Tiger He

Data Science Capstone Project with Accenture

Background

Conversations can be a useful source of learning about practically any topic. No matter it is a phone call between a customer and an agent, or a simple daily dialogue between friends, important information can be revealed through studying the structure and organization of the talk, and often far more valuable than one might expect. Hence, we have analyzed two different datasets, one focusing on daily dialogues while the other presenting real-life call transcripts.

Goal

We mainly focused on detecting the act of dialogues that may provide insights into further analysis. As a crucial step toward understanding spontaneous speech and more, dialogue act detection can further discover a user's intention. In the process of identifying patterns in real-life communication systems, we aim to find out what communication partners actually do, rather than what they think or say they do.

Method

We have six different classes for the label: State-Non-Opinion, Statement-Opinion, Acknowledge, Apology, Question, and Other. To classify the datasets into the 6 given classes, we started from feature engineering and ran multiple models. We also used deep learning models to generate machine learning features (sentiment and emotion scores). 4 different tree-based models are then used with their results compared.

Dataset and Preprocessing

Input Dialogue Data: 1000 samples

Preprocessing (part1)

- remove punctuations
- remove stop words
- split participants A/B

…

Feature Engineering(part 1)
Extracted 21 features:

- # of words ratios of A vs B
- # of negations
- freq of pronouns
- # of sentences in conversation
- # of past tense verbs
- # of future tense verbs

…

Preprocessing (part2)

- stemming/lemmatization
- tokenization
- target encoding

Feature Engineering(part 2)
Generated features using deep learning pretrained models:

- scores comes from softmax layer
- sentiment scores
  - roBERTa-base model
  - Pretrained model on English language using a masked language modeling (MLM) objective.
- emotion scores
  - distillBERT model
  - small, fast, cheap and light Transformer model

Model Training and Validation

Model Comparison

- Random Forest
- XGBoost
- LightGBM
- CatGBM

…

Hyperparameter Tuning with Cross Validation

- Learning rate
- Min_sample_split
- N_estimators
- Loss function

…

Choose Best Model (Metrics: F1 score)

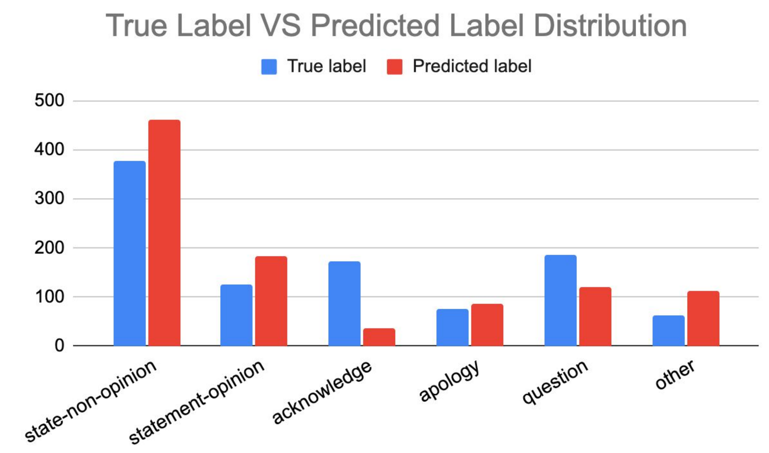
F1-score	state-non-opinion	statement-opinion	acknowledge	apology	question	other
LightGBM	0.71	0.69	0.21	0.81	0.69	0.63
XGBoost	0.68	0.56	0.32	0.62	0.28	0.17
RF	0.61	0.32	0.28	0.64	0.33	0.56
CatGBM	0.65	0.33	0.17	0.75	0.68	0.61

Final Training Using best hyperparameters on all data samples using final model
(1000 samples in total)

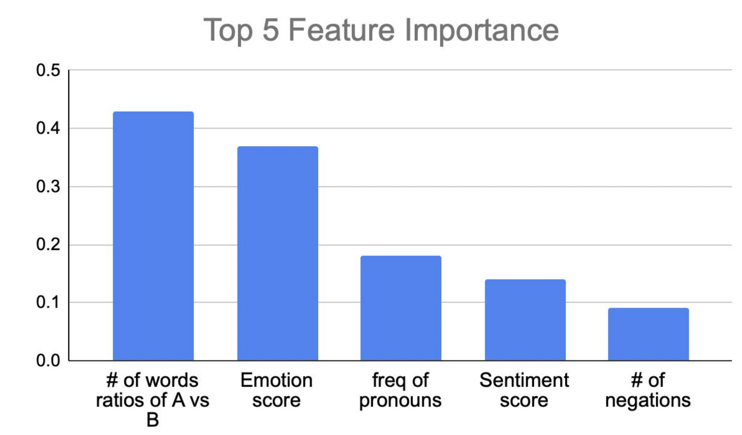
Model Deploying

- **Final Model Architecture**
- **Feature Importance**
- **Exhausted list of acts**

True labels **Predicted labels**



Feature Importance



Conclusion

We evaluated the performance of several models on dialogue acts prediction task based on our 6 label classes. Specifically, we have chosen tree models over Deep Learning models based on how tree models have their own advantages, which are faster and have lower training cost. In addition, tree-based models allow us to derive feature importance, revealing more information and giving better model interpretation.

In the process of feature engineering, we have compared different features and analyzed how they affect the classification of dialogue acts as well as which of them are important to our final model and how. The top 5 features are discussed in more details.

In the future, we plan to include more types of models beyond the tree models, like CNN, RNN and other combined models. We would also consider evaluating predicting results with and without contextual information to analyze if the dialogue act classes are more general and has less dependency on the context or the dependency can be very strong. We also consider fine-tuning our own high-quality sentence transformers.

References

[1] N Reithinger, M Klesen. " Dialogue act classification using language models" 5th European Conference on Speech Communication and Technology Rhodes, Greece, September 22-25 (1997)

[2] Harry Bunt. 2006. Dimensions in Dialogue Act Annotation. In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy. European Language Resources Association (ELRA) (2006)