Research

The goal of this project is to build a text classification model that can predict the action, object, and location from a given command. The first step was to research and gather data on the problem domain. After reviewing the problem domain, it was determined that the problem can be modeled as a multi-class classification problem. The dataset for this problem was collected from various sources, and it consists of a set of commands and their corresponding action, object, and location labels.

Data Preparation

The next step was to prepare the data for training. The data was loaded into a Pandas dataframe and cleaned by removing any missing values. The text data was preprocessed using the following steps:

Lowercasing

Removing punctuation

Tokenization

After preprocessing the text data, the labels were combined into a single column, separated by underscores. The data was then split into training and validation sets.

Model Training

The model used for this project is a Support Vector Machine (SVM) with a linear kernel. The text data was transformed into feature vectors using the TfidfVectorizer from scikit-learn. The labels were encoded using LabelEncoder from scikit-learn.

The hyperparameters for the model were selected through a combination of manual tuning and grid search using cross-validation. The final hyperparameters were:

C: 1.0

kernel: 'linear'

class_weight: 'balanced'

The model was trained on the training data and evaluated on the validation data using accuracy, precision, recall, and F1-score metrics.

Model Deployment

The final step was to deploy the model for inference. The model was serialized and saved to disk along with the encoder and vectorizer. A CPU inferencing script was created that loads the saved model and makes predictions on new text data. The script can be configured using a YAML file that specifies the parameters for the vectorizer and the model. The script was also tested on the test data to ensure that it is working correctly.

Conclusion

In conclusion, this project successfully addressed the problem of text classification for command prediction. The SVM model with a linear kernel achieved a high F1-score on the validation data and was deployed for inferencing using a CPU inferencing script. This project demonstrates the importance of data preparation, model selection, hyperparameter tuning, and model deployment in creating effective text classification models.