Harish Ram, Sisi Zhang, Zhuohan Yu

Group 5 - Final Project Proposal

Background
GLUE is a website containing numerous datasets and their corresponding tasks which typically consist of collecting predictions based on training a model on a training, development and test set which are all provided on the GLUE website. Each dataset comes with a specific task that requires collecting the best F1/accuracy score, Mathew's Corr or Pearson's Corr. In this project, our group will be completing all the tasks on GLUE, reaching the baseline value and fine-tuning our models to exceed the benchmark.

Data Source
The datasets on GLUE come from multiple sources and concern varying topics all concerning natural language processing. The overall tasks that GLUE contains involve either single sentence tasks, similarity and paraphrase tasks and inference tasks. The single sentence tasks use two datasets; one involves deciphering whether a sentence is grammatically acceptable and the other using sentiment analysis on a set of movie reviews.

The similarity and paraphrase tasks are done on several datasets, the first being a corpus of sentence pairs from online news sources in which the task is to determine semantic similarity. The next dataset is a question pair corpus coming from the online forum website, Quora to check if two questions are similar. The final dataset for this section is a collection of sentence pairs pulled from news sources and media captions with a corresponding similarity score from 1-5.

The first dataset used for the inference task is in the form of sentence pairs (the first sentence is a premise and the second is a hypothesis) sourced from speech, fiction and government reports. The goal is to see whether the premise agrees, disagrees or has no relation to the hypothesis. The next dataset is a question-answer pair in which the answer sentence must be extracted from a corresponding paragraph. The answer may or may not be an accurate answer, so the task is to evaluate its accuracy. The following dataset comes from Wikipedia and news text with the task being to determine whether the meaning or context of a sentence fragment can be inferred from the other. The last dataset for this GLUE task is sentence-word pair dataset in which the word is a pronoun that refers to a word in the sentence. In order to complete this task, we must first understand the context of the sentence and accurately determine where the pronoun should be placed for the context of the sentence to still make sense.

Schedule

11.5 (1 week)	Group proposal and research
11.19 (2 weeks)	Preprocessing (tokenization, create features)
11.26 (1 week)	Modeling (try different models) and improve score
12.3 (1 week)	Report write up
12.9 (1 week)	Final project presentation and presentation

Harish Ram, Sisi Zhang, Zhuohan Yu

Dataset	Tasks	Metrics
The Corpus of Linguistic Acceptability(CoLA)	Text Classification (Predict if the sentences are grammatically correct or not)	Matthew's Corr
The Stanford Sentiment Treebank(SST-2)	Sentiment Analysis (Predict if the move reviews has positive or negative sentiment)	Accuracy
Microsoft Research Paraphrase Corpus(MRPC)	Semantic Textual Similarity (To identify if two sentences are paraphrases of each other)	F1/Accuracy
Semantic Textual Similarity Benchmark(STS-B)	Semantic Textual Similarity (Each pair with a similarity score from 1 to 5)	Pearson-Spearman Corr
Quora Question Pairs(QQP)	Semantic Textual Similarity (Determine whether a pair of questions are semantically equivalent)	F1/Accuracy
MultiNLI Matched(MNLI)	Inference (Determine whether hypothesis relates to premise - in-domain)	Accuracy
MultiNLI Midmatched	Inference (Determine whether hypothesis relates to premise - Cross-domain)	Accuracy
Question NLI(QNLI)	Question Answering (Answer generated from paragraph - evaluate answer accuracy)	Accuracy
Recognizing Textual Entailment(RTE)	Textual entailment (Determine if context of one sentence can be inferred from another)	Accuracy
Winogard NLI(WNLI)	Inference (Demonstrate pronoun placement given pronoun and sentence containing referents)	Accuracy
Diagnostics Main	Textual Entailment (Determine if sentence pairs entail, contradict or do not describe each other)	Matthew's Corr

Data Source URL:
https://gluebenchmark.com/tasks

Harish Ram, Sisi Zhang, Zhuohan Yu

Github URL:
https://github.com/HarishRam10/Final-Project5.git