XCS 224U

Bake-off Report

Sri Vardhamanan (Thanks to all other CFs & CDs)

Task Description

Dataset

1. DynaSent Round 1 & Round 2
2. Stanford Sentiment Treebank (SST)

Design Choices

1. Classifier structure
2. Feature extraction
   a. Model Choices
   b. Pooling Method
3. Dataset Preparation

Evaluation

1. Macro F1
2. Test Data
   a. DynaSent R1 and R2 (Test)
   b. SST (Test)
   c. Mystery Examples

Top Distinguishing Factor

Strategies that positively impacted the final performance

Starting with good LMs

Capable LMs: Electra, Roberta
Seq Length: 128
Representation: Avg. Pooling
Low Learning rate: 1e-5 -> 5e-5
Better trains: Early stopping

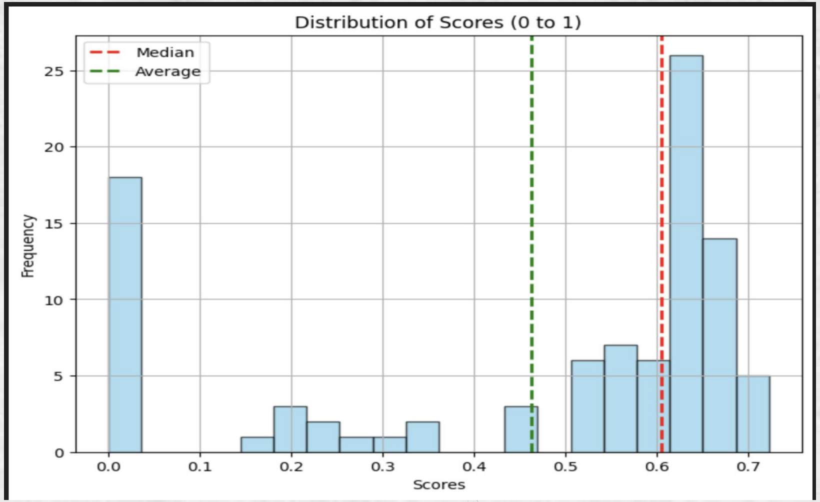
Effective (Pre) Classifier

Expressiveness: Non-linear Layers, wider & deep network
Regularization: Dropouts

Dataset Tuning

Maximize: combine DynaSent R1, R2, and SST
More Data: Amazon reviews, SLIDE
Balanced Classes: resampling

Score Distribution



25th Percentile → 0.24

50th Percentile → 0.60

75th Percentile → 0.64

Overachieving Systems

01

Venetis Pallikaras

Macro f1 → 0.721

Differentiators - Meticulous Training

- ❑ ***Maximize Data Usage:*** Combine DynaSent R1, R2, and SST
- ❑ ***Experiment Array of LMs:*** Bert, DistilBERT, RoBERTA, Electra (Best performing), DeBERTA
- ❑ ***Better Training:*** Warmup steps, Weight decay, Low LR, FP16, Early Stopping ★

Overachieving Systems

02

Santiago Ibanez Lopez

Macro f1 → 0.707

Differentiators - Expressive System

❑ ***Maximize Data Usage:*** Combine DynaSent R1, R2, and SST

❑ ***Maximize Learning:*** Max seq len 512 -> 128, higher batch size, early stopping

❑ ***Expressive Representation:*** Use of all hidden states instead of mean/max pooling ★

❑ ***Better Classifier:*** 2 FC + non-linearity ★

❑ ***More Capable LM:*** RoBERTA

Overachieving Systems

03

Sugi Venugeethan

Differentiators - A step forward in every direction

Macro f1 → 0.692

❑ ***More Capable LM:*** BERT, RoBERTA, and DistilBERT + spacy.
❑ ***Maximize Data Usage:*** Combine DynaSent R1, R2, and SST
❑ ***Data fairness:*** Re-sampling for balanced class labels
❑ ***Better Classifier:*** Dropouts, Non-linearity
❑ ***Better Representation:*** Pooler output

Interesting Approaches

Marcello Esposito	Ensemble of 4 different models.
Pierre Cadman	Amazon reviews as additional sentiment dataset.
Kelvin Kakugawa	Data augmentation using C4 dataset, BM25 & Llama2.
Milan Hejtmanek	Data Centric Approach: intensive data cleaning- transform emoji, handling foreign language text, remove contaminated texts, etc.
Ankit Kumar Patel	SLIDE as additional sentiment dataset and Attention on last hidden state outputs
Caroline Silva	Soft-prompting & LoRA on BLOOM
Hamilton Link	Part-of-Speech Count Vectorizer to train with decision tree random forest
Igor Khomyakov	Nearest Neighbour Classifier with BERT Embeddings
Yogesh Luthra	Evaluation on various token representation methods: pooler output, mean across token representation, and masked mean across token representation. Wider & Deeper classifier with ReLU and masked mean across token representation.
Asad Ezazi	Fine-tuning while freezing different set of BERT layers: first 2, last 2, and all hidden layers

Awesome Work, Everyone!