Natural Language Inference

Shira Rozenthal and Peter Sapountzis

Causal Inference & NLI

Causal inference is the process of determining if a causal relationship exists between two events.

In NLI, the relationship between a premise and a hypothesis can be classified into one of 3 categories:

1. **Entailment**
2. **Neutrality**
3. **Contradiction**

Textual Entailment

For example, the premise “Peter and Shira are presenting their NLP project”

Entails that Peter and Shira are on the zoom call

Contradicts that Peter and Shira are sleeping through class

Is ***neutral*** on whether they are having fun

…we’re having fun we promise!

MNLI

The MultiNLI (MNLI) dataset, sourced from the Open American National Corpus, encompasses both spoken and written premise/hypothesis pairs across 10 genres.

- These genres include transcriptions of face-to-face conversations, government documents, the 9/11 Commission report, articles from Slate Magazine, telephone conversations, etc.

- A distinctive aspect of MNLI is that only half of the genres are included in the training set, with the remaining unseen genres serving to test a model's ability to generalize to new text sources.

premise string · *lengths* 5 1.82k	premise_binary_parse string · *lengths* 6 3.43k	premise_parse string · *lengths* 23 6.45k	hypothesis string · *lengths* 1 393	hypothesis_binary_parse string · *lengths* 1 669	hypothesis_parse string · *lengths* 19 1.15k	genre string · *classes* 5 values	label class label 3 classes
The other men shuffled.	((The (other men)) (shuffled .))	(ROOT (S (NP (DT The) (JJ other) (NNS men)) (VP (VBD shuffled)) (.…	The other men were shuffled around.	((The (other men)) ((were (shuffled…	(ROOT (S (NP (DT The) (JJ other) (NNS men))…	fiction	0 entailment
States must show reasonable…	(States ((must ((show ((…	(ROOT (S (NP (NNPS States)) (VP (MD must) (VP (VB show) (NP (NP (JJ…	Itis not necessary for there to be…	(((Itis (not necessary)) (for (…	(ROOT (S (VP (VB Itis) (ADJP (RB not) (JJ…	government	2 contradiction
well it's been very interesting	(well (it ('s (been (very…	(ROOT (ADJP (RB well) (SBAR (S (NP (PRP it)) (VP (VBZ 's) (VP (VBN…	It has been very intriguing.	(It ((has (been (very intriguing)))…	(ROOT (S (NP (PRP It)) (VP (VBZ has) (VP (VBN…	telephone	0 entailment
He started slowly back to the…	(He (((started (slowly back)) (…	(ROOT (S (NP (PRP He)) (VP (VBD started) (ADVP (RB slowly) (RB…	He returned slowly to the bunkhouse.	(He (((returned slowly) (to (the…	(ROOT (S (NP (PRP He)) (VP (VBD returned) (ADV…	fiction	0 entailment
and it's it's quite a bit i…	(and (it ('s (((((it (('s (…	(ROOT (PRN (CC and) (S (NP (PRP it)) (VP (VBZ 's) (S (S (NP (PRP it)) (V…	I do not know exactly where the…	(I (((do not) (know ((exactly wher…	(ROOT (S (NP (PRP I)) (VP (VBP do) (RB not)…	telephone	1 neutral
They're made from a secret recipe…	(They (('re (made (from ((a …	(ROOT (S (NP (PRP They)) (VP (VBP 're) (VP (VBN made) (PP (IN from)…	The recipe passed down from…	(((The recipe) ((passed down) (from …	(ROOT (S (NP (NP (DT The) (NN recipe)) (VP…	travel	2 contradiction
yeah well you're a student right	(yeah (well (you ('re (a (studen…	(ROOT (S (VP (VB yeah) (S (ADVP (RB well)) (NP (PRP you)) (VP (VBP 're)…	Well you're a mechanics student…	(Well (you (('re (a (mechanics (stude…	(ROOT (S (ADVP (RB Well)) (NP (PRP you))…	telephone	1 neutral

BERT 

- We are using the mDeBERTa-v3-base-xnli-multilingual-nli-2mil7 model, developed by Microsoft, for our project.

- Built on the mDeBERTa-v3-base architecture and pre-trained on the CC100 dataset, this model supports natural language inference (NLI) in 100 languages and is ideal for multilingual zero-shot classification tasks.

- We are evaluating model performance with validation accuracy.

Training Our Model

- **Determining Optimal Subset to Train**: Given that our model is already pre-trained on some natural language, we wanted to determine where our validation accuracy begins to plateau in order to optimize our runtime without sacrificing performance.

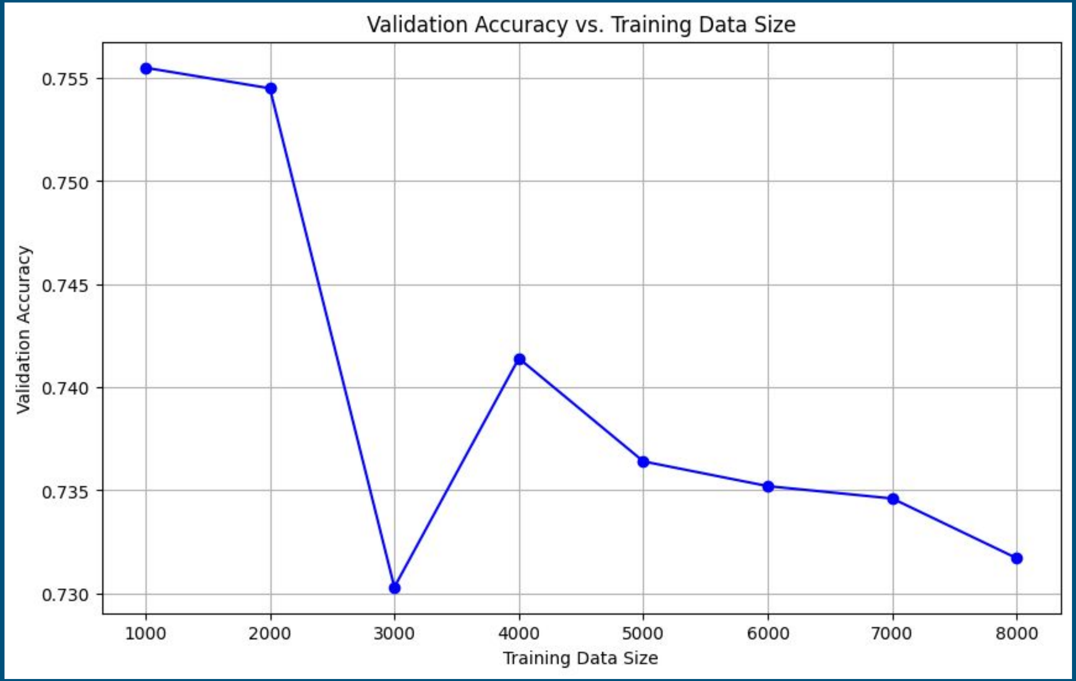
- **Gradient Clipping**: This prevents the gradients from exploding, which is crucial for maintaining stable training dynamics.

- **Learning Rate Scheduler**: Adjusts the learning rate throughout training, which can help in converging to a better minimum.

```
Epoch 1: Average Loss: 0.37268252827463644
Epoch 2: Average Loss: 0.08028289416642298
Epoch 3: Average Loss: 0.017550210266130135
Epoch 4: Average Loss: 0.010786756653004208
```

```
Validation Accuracy: 85.26%
```

Demo


Validation Accuracy vs. Training Data Size
Validation Accuracy
0.755
0.750
0.745
0.740
0.735
0.730
1000
2000
3000
4000
5000
6000
7000
8000
Training Data Size