|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **SICK**  **Entailment** | | **SICK**  **Relatedness** | **MSRP** | **AI2-8grade** | | **QQP** | |
| Transformer Siamese Encoder | Head | | 83.17 | 72.75/0.5129 | 75.90 | | 75.29 | | 80.99 |
| Branch | | 82.63 | 73.04/0.5042 |  | | 74.78 | |  |
| Transformer Encoder Once Non Siamese | Head | | 84.80 | 72.78/0.5127 | **76.40** | | **77.29** | |  |
| Branch | | **85.22** | 73.18/0.4969 | 75.29 | | 74.20 | | 84.07 |
| InferSent | | 84.62 | | **85.63**/0.2732 | 74.46 | **77.29** | | 85.28 | |
| LSTM | | 76.80 | | 82.91/0.3244 | 70.74 | 76.97 | | 82.48 | |
| BiLSTM Projection Layer | | **85.22** | | 80.37/0.3667 | 74.24 | 74.88 | |  | |
| BiGRU Last Encoder | | 81.47 | | 83.17/0.3147 | 70.46 | 74.76 | | 82.44 | |
| Inner Attention | | 72.01 | | 78.63/0.3944 | 69.74 | 74.77 | | 82.44 | |
| ConvNet Encoder | | 83.82 | | 85.20/0.2806 | 73.96 | 75.43/77.33 | |  | |

GRU

mse: 0.3178815245628357

pearson: 0.8313440680503845

loss: 0.10587805443304803

Inner Attn NAACL

mse: 0.5501508712768555

pearson: 0.6981591582298279

loss: 0.14438188115468653

Inner Attn YANG

mse: 0.4547271430492401

pearson: 0.7556540966033936

loss: 0.1357176222002785

Blstm proj

mse: 0.36677098274230957

pearson: 0.8036734461784363

loss: 0.11628564453874897

InferSent

mse: 0.27321305871009827

pearson: 0.8563084602355957

loss: 0.09872134307305636

BiGRU

mse: 0.3147765100002289

pearson: 0.8317111730575562

loss: 0.10562648809803579

Inner Attn MILA

mse: 0.39446479082107544

pearson: 0.7863677144050598

loss: 0.12452546356118095

LSTM

mse: 0.3403347134590149

pearson: 0.8179712891578674

loss: 0.11060604058130094

transformer 5 mse: 0.6258545517921448

pearson: 0.6331384778022766

loss: 0.15608961344058625

transformer 2 mse: 0.5677379369735718

pearson: 0.6830798387527466

loss: 0.14873846943552774

transformer 5 kappa mse: 0.5721760988235474

pearson: 0.6659924387931824

loss: 0.14977495014056585

Infersent 5 kappa mse: 0.3950158655643463

pearson: 0.7846205830574036

loss: 0.12368686345961324

transformer 2 ED mse: 0.5035095810890198

pearson: 0.7318350672721863

loss: 0.14692964898174737

lstm manhattan mse: 0.37855416536331177

pearson: 0.798143208026886

loss: 0.3785541827405472

transformer manhattan ED mse: 0.6497915387153625

pearson: 0.6624051332473755

loss: 0.6497916982935532

**December:**

**Background Research:** Done some background research on multi-lingual word embeddings (MUSE), existing sentence encoding architectures like InferSent and Transformer, phrase extraction from text (AutoPhrase), word ontologies, sentence similarity evaluation metric (biText similarity, arccos similarity, manhattan distance etc.).

**Dataset Preprocessing:** Done some preprocessing stuffs like stop words removal, noise reduction, sentence alignment, translation, dictionary preparation, performance metric selection etc.

**Model Preparation:** Evaluating sentence pairs based on the currently available pre-trained models, develop a multi-lingual sentence similarity model from InferSent and modified cross attention Transformer, training the word embedding model on the actual corpus instead of using a pre-trained one (Currently working).

**January:**

- Performed experiments using Infersent’s English-French model over cic dataset. Checked the generalization ability of the model by testing it over European Parliament dataset.

- Made an English-French version of MSRP and SICK dataset using Google’s translation API.

- Did experiments over these datasets using the Transfer learning mechanism.

- Currently working on extracting the phrase embedding using Tree structured models.