COMPARATIVE REPORT

Papers:

1. Pre-trained Language Model Representations For Language Generation, LINK: https://www.aclweb.org/anthology/N19-1409.pdf
2. PEGASUS: Pre-training With Extracted Gap-sentences For Abstractive Summarization,
   LINK: http://proceedings.mlr.press/v119/zhang20ae/zhang20ae.pdf
3. Controllable Abstractive Sentence Summarization With Guiding Entities,
   LINK: https://www.aclweb.org/anthology/2020.coling-main.497.pdf

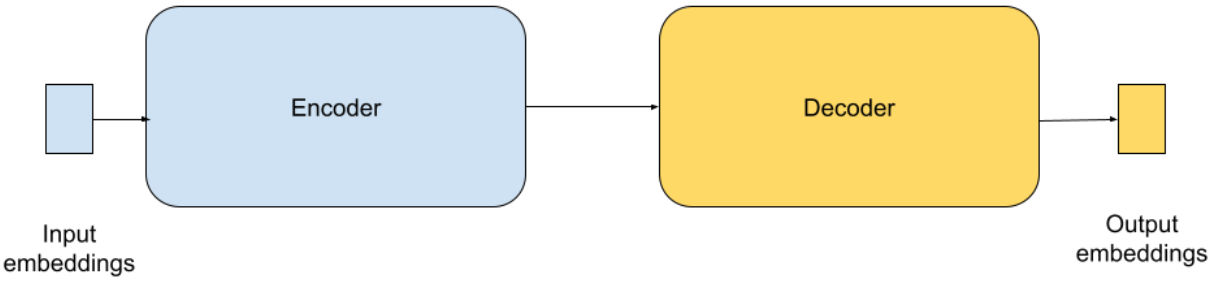
S.No	Name	Roll Number
1	Anurag Saraswat	M20CS066
2	Parsa Revanth	M20CS058

Objectives

- In paper 1, Pre-trained Language Model Representations For Language Generation, the objective is to use pre-trained representations on a sequence to sequence encoder-decoder architecture and evaluate the model on tasks like machine translation and abstractive summarization
- In paper 2, PEGASUS, the objective is to use a pre-trained encoder-decoder model for abstractive summarization by using a technique gap sentence generation
- In paper 3, Controllable Abstractive Sentence Summarization With Guiding Entities, the objective is to use the entity information along with the sentence representations and generate the summaries with entities in the output summaries

Architecture

- All the three papers use an encoder-decoder architecture
- The baseline sequence to sequence encoder-decoder architecture is,

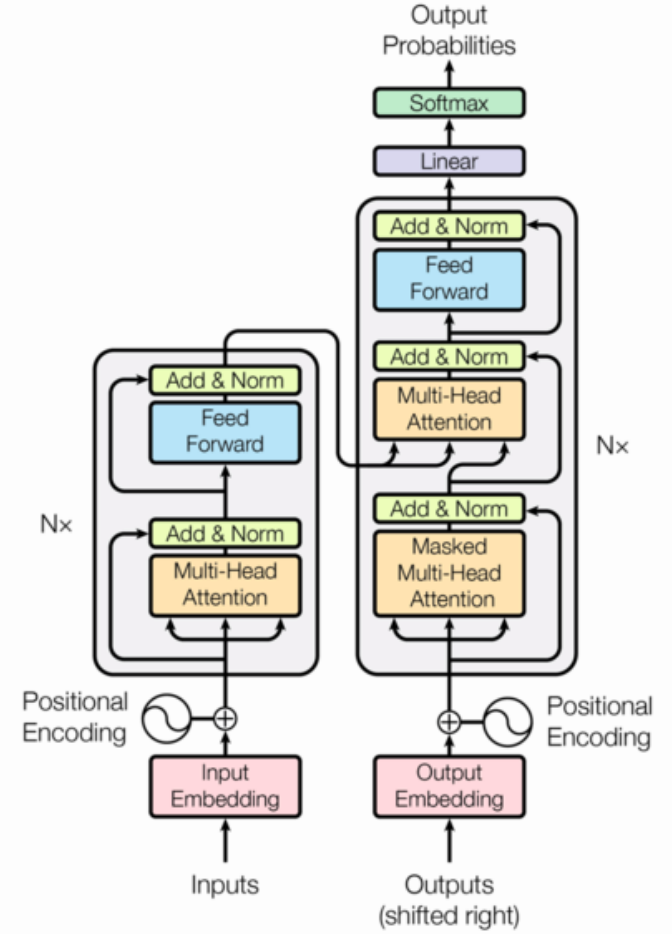


- In paper 1, the authors defined 5 different architectures all use transformers as encoders or decoders and they made these architectures depending on the type of pretraining and the sharing of the embeddings in the encoder and decoder
- In paper 2, the authors also used the transformers as both encoders and decoders and pretarin in huge corpus and use pre trained architecture for downstream tasks like abstractive summarization
- In paper 3, the authors used pretrained BERT embeddings with Bi-LSTM as an encoder and two LSTMs which creates the representations for entities by

looking the sequence from the left to right by one LSTM and right to left by other LSTM

- BERT is also based on the transformer architecture only

- The transformer architecture is,



- The transformer architecture has a multihead attention which makes it a different and also prominent in the area
- The main difference between paper 1's best model (SRC- ELMO + SHREMD) and model in paper 2 is that there are shared embeddings in the decoder of the paper 1's model and paper 2's model doesn't share any embeddings

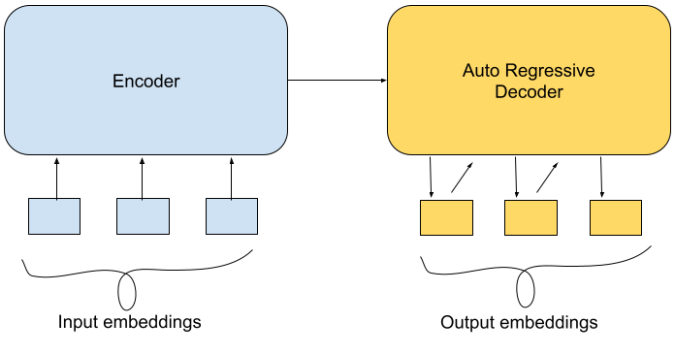


Fig 1. Shared embedding at decoder architecture

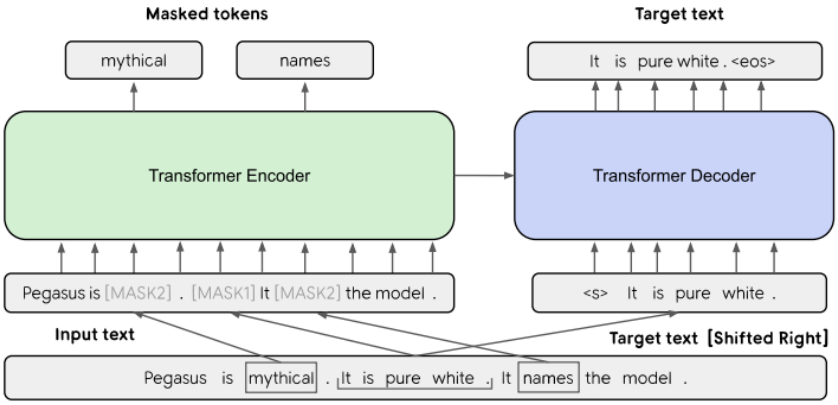


Fig 2. Encoder-decoder architecture used in PEGASUS

Evaluation metrics

- The evaluation metrics defined in the paper are ROUGE, BLUE, Informativeness Metric, Human evaluation
- ROUGE is the Recall-Oriented Understudy for Gisting Evaluation, there are three variants namely used in the papers which are Rogue-1, Rouge-2, Rouge-L
    1. Rouge-1: This will check the unigram overlaps in the target sentence to that of the reference sentence given
    2. Rouge-2: This will check the bigram overlaps in the target sentence to that of the reference sentence given
    3. Rouge-L: This will check the longest common subsequence in the target sentence to that of the reference sentence
- BLUE (bilingual evaluation understudy) scores are calculated for individual translated segments—generally sentences—by comparing them with a set of

good quality reference translations. Those scores are then averaged over the whole corpus to reach an estimate of the translation's overall quality.

- Informativeness Metric : The metric can evaluate not only the relevance between model outputs and manual references, but also the extra and omissive information in generated summaries. Compared to existing lexicon based metrics, it is more human-like and helps to analyze results more comprehensively
- Human evaluation : In this scheme randomly choose samples and generate summary using model and ask different evaluators to rate summaries

Datasets

- In paper 1, CNN-DailyMail news dataset is used for abstractive summarization task and WMT'18 datasets are used for English to German and English to Turkish machine translation task
- In paper 2, C4 and Hugenews datasets are used to train the encoder-decoder model and also used 12 other datasets for evaluation which are XSum, CNN/DailyMail, NEWSROOM, Multi-News, Gigaword, arXiv, PubMed, BIGPATENT, WikiHow, Reddit TIFU, AESLC, BillSum
- In paper3, English Gigaword corpus and DUC-2004 datasets are used to evaluate the model