

# PEGASUS Summarization

Pre-training with Extracted Gap-sentences  
for Abstractive Summarization



520V0015 – Trung Nguyen  
520K0163 – Trong Nghia

# Pegasus Overview

- Pegasus' pretraining task is intentionally similar to summarization: important sentences are removed/masked from an input document and are generated together as one output sequence from the remaining sentences, similar to an extractive summary.
- Pegasus achieves SOTA summarization performance on all 12 downstream tasks, as measured by ROUGE and human eval.
- The base model has 16 encoders-16 decoders

# PEGASUS Variants

- **PEGASUS-fine-tuned models:** have been fine-tuned on specific NLP tasks
- **PEGASUS-PP:** is designed specifically for the task of paraphrasing
- **PEGASUS-Multi:** extends the PEGASUS architecture to multiple languages
- **PEGASUS-XL:** This larger model achieved state-of-the-art performance on several summarization benchmarks,

# Transformer

Before Transformer was announced, most natural language processing tasks, especially machine translation (Machine Translation) used Recurrent Neural Networks (RNNs) architecture.

The weakness of this method is that it is difficult to catch the long dependence between words in the sentence and the training speed is slow due to sequential input processing.

# Self-supervised learning

Self-supervised learning is a type of machine learning where a model learns to predict certain features or properties of the input data, without the need for labeled data. The idea behind self-supervised learning is to use the vast amount of unlabeled data that is available to train models that can learn useful representations of the data.

# Pegasus Self-supervised Objective

The hypothesis is that the closer the pre-training self-supervised objective is to the final down-stream task, the better the fine-tuning performance.

The base architecture of PEGASUS is a standard Transformer encoder-decoder. Both GSG and MLM are applied simultaneously to this example as pre-training objectives. Originally there are three sentences. One sentence is masked with [MASK1] and used as target generation text (GSG). The other two sentences remain in the input, but some tokens are randomly masked by [MASK2] (MLM).

# Evaluation

- ROUGE (Recall-Oriented Understudy for Gisting Evaluation): computes the similarity of two texts by computing n-gram overlaps using a score from 0 to 100 (ROUGE-1, ROUGE-2, and ROUGE-L are three common variants).
- BLUE (Bilingual Evaluation Understudy): compares each translation segment with a set of reference translations with good translation quality and calculates each segment score then estimates the overall quality of the translation.