PEGASUS : Pre-training with Extracted Gap-sentences for Abstractive Summarization

Zhang, Jingqing, et al, *Int. Conf. on Machine Learning. PMLR, 2020*

Seungone Kim

Department of Computer Science

Yonsei University

louisdebroglie@yonsei.ac.kr

2021.07.21

연세대학교 YONSEI UNIVERSITY

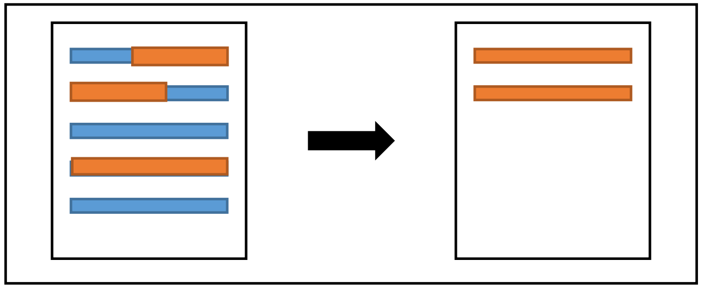
SOFT COMPUTING LABORATORY

Outline

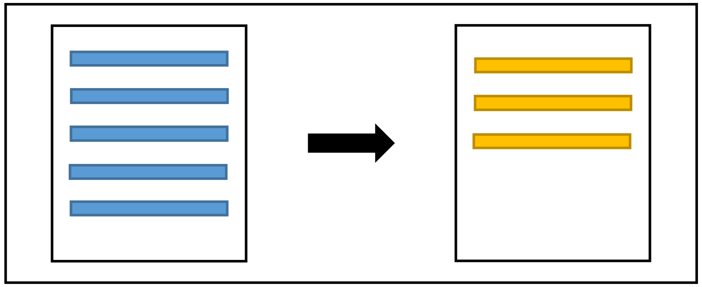
- 문제 정의
  - Extractive, Abstractive Summarization
  - ROUGE Metric
  - Datasets used for fine-tuning in Summarization tasks
- 관련 연구
  - Limitations of existing methods and models
- 제안하는 방법
  - Gap Sentences Generation(GSG)
- 실험 결과
- 결론

문제 정의

- Extractive Summarization
  - Binary Classification task with labels indicating whether a text span (typically a sentence) should be included in the summary
  - *Liu et al., "Text Summarization with Pretrained Encoders" EMNLP-IJCNLP 2019*
  - Picking top-k sentences as the representative summary



- Abstractive Summarization
  - Task of generating a short and concise summary that captures the salient ideas of the source text, which could potentially contain new phrases and sentences that may not appear in source text
  - *Liu et al., "Generative Adversarial Network for Abstractive Text Summarization" AAAI 2018*
  - Generating novel sentences by rephrasing or using new words



문제 정의

- ROUGE Score (Recall-Oriented Understudy for Gisting Evaluation)
  - Goal of Summarization : Shorten source document with including principal information and the summarized result should be linguistically fluent
  - A metric used to evaluate how well summarization is performed
  - *Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." Text summarization branches out. 2004.*
  - Based on N-gram, compare the summarization and the original text, and count how many subset matches between the two
  - *May not be the perfect measure to determine if the given summary is good or not**

- $$Rouge-N=\frac{\sum_{S\in\{Reference\ Summaries\}}\sum_{gram_n\in S}Count_{match}(gram_n)}{\sum_{S\in\{Reference\ Summaries\}}\sum_{gram_n\in S}Count(gram_n)}$$

문제 정의

- CNN / DailyMail news highlights dataset
  - *Chen, Danqi, Jason Bolton, and Christopher D. Manning. "A Thorough Examination of the CNN/Daily Mail Reading Comprehension Task." Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016.*
  - CNN Dataset Size : 90,266 / 1,220 / 1,093 documents for training / validation / testing (198MB)
  - DailyMail Dataset Size : 196,961 / 12,148 / 10,397 documents for training / validation / testing (460MB)
  - Contains news articles and associated highlights
  - Extractive Summarization Dataset

- New York Times Annotated Corpus
  - *Durrett, Greg, Taylor Berg-Kirkpatrick, and Dan Klein. "Learning-Based Single-Document Summarization with Compression and Anaphoricity Constraints." Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016.*
  - Size : 100,834 / 9,706 documents for training / test examples (Use 4000 documents from training example for validation set)
  - Contains articles with abstractive summaries
  - Extractive Summarization Dataset

문제 정의

- English Gigaword
    - *Napoles, Courtney, Matthew R. Gormley, and Benjamin Van Durme. "Annotated gigaword." Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction (AKBC-WEKEX). 2012.*
    - Size : 3,803,957 / 189,651 / 1,951 documents for training, validation, testing (551MB)
    - Contains news articles and associated headlines
    - Abstractive Summarization Dataset

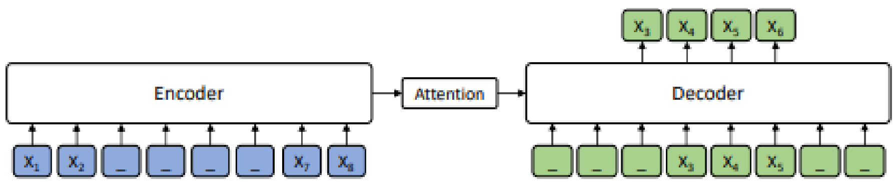
- Wikihow
    - *Koupaee, Mahnaz, and William Yang Wang. "Wikihow: A large scale text summarization dataset." arXiv preprint arXiv:1810.09305 (2018).*
    - Size : 157,252 / 5,599 / 5,577 documents for training, validation, testing (5MB)
    - Contains articles with abstractive summaries
    - Abstractive Summarization Dataset

문제 정의

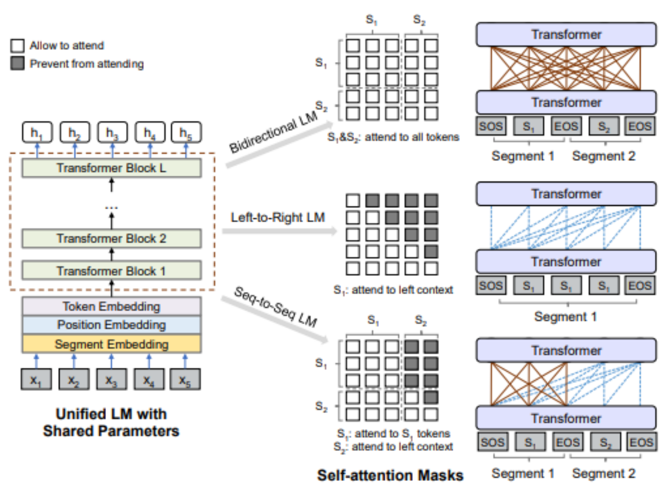
- XSum
    - *Narayan, Shashi, Shay B. Cohen, and Mirella Lapata. "Don’t Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization." Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018.*
    - Size : 204,045 / 11,332 / 11,334 documents for training, validation, testing
    - Abstractive Summarization Dataset

관련 연구

- MASS
  - *Song, Kaitao, et al. "MASS: Masked Sequence to Sequence Pre-training for Language Generation." International Conference on Machine Learning. PMLR, 2019.*
  - Reconstructs a sentence fragment given the remaining part of the sentence
  - A single sentence fragment was randomly selected

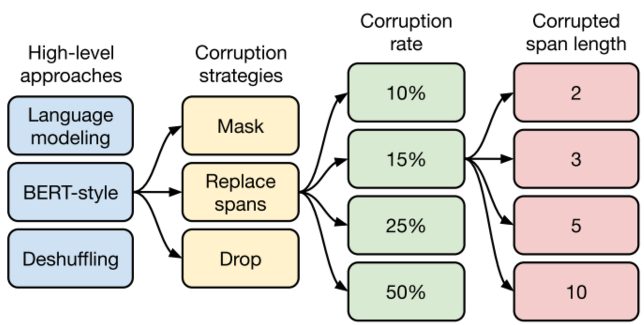


- UniLM
  - *Dong, Li, et al. "Unified language model pre-training for natural language understanding and generation." Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019.*
  - Jointly trains on Unidirectional, Bidirectional(word level), Sequence-to-Sequence(word level) prediction

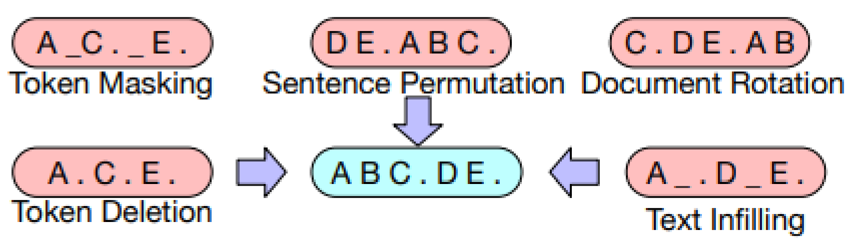


관련 연구

- T5
  - *Raffel, Colin, et al. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." Journal of Machine Learning Research 21 (2020): 1-67.*
  - Pre-trains with randomly corrupted text spans of varying mask ratios and sizes of spans



- BART
  - *Lewis, Mike, et al. "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.*
  - Text Infilling which used single mask tokens to mask random sampled spans of text



관련 연구

- Pre-training Objectives tailored for abstractive text summarization not explored
  - Models were fine-tuned just for summarization after pre-training
  - More efficient to come up with Pre-training Objective for abstractive summarization

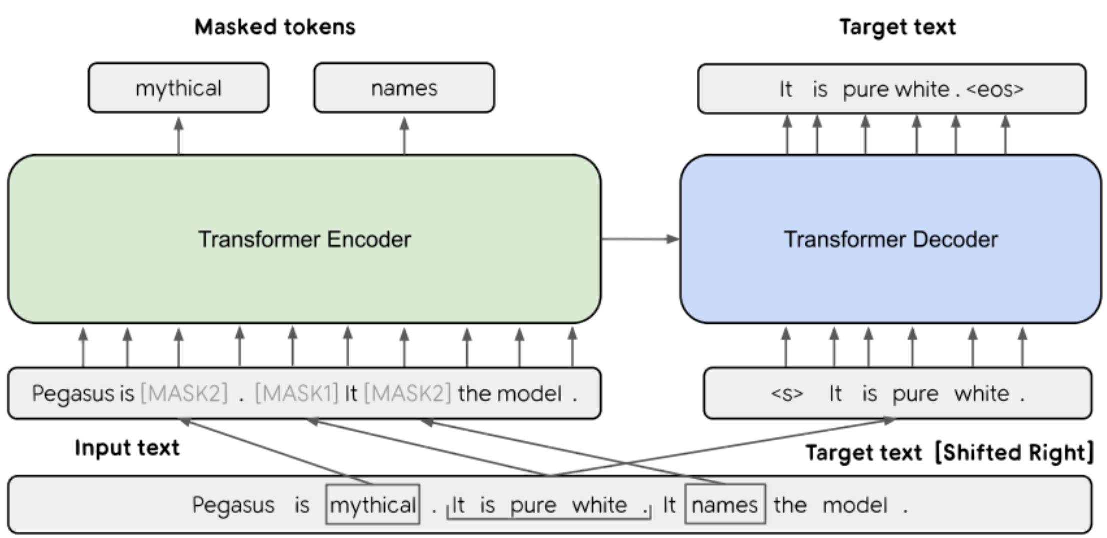
- Lack of systematic evaluation across diverse domains
  - PEGASUS was evaluated on 12 downstream summarization tasks
  - News, Science, Stories, Instructions, Emails, Patents, Legislative Bills

제안하는 방법

- Pretraining PEGASUS for Abstractive Summarization
  - PEGASUS masks multiple whole sentences rather than smaller continuous text spans
  - Deterministically choose sentences based on importance, rather than randomly
  - Generate the masked sentences as a single output sequence
  - Inspired by recent success in masking words and contiguous spans (SpanBERT, T5)
  - Focus entirely on downstream summarization tasks and do not evaluate on NLU classification tasks

제안하는 방법

- (Propose) Gap Sentences Generation
  - Remove/Mask important whole sentences from input document and generate together as one output sequence from remaining sentences
  - Choosing putatively(추정) important sentences outperforms lead or randomly selected ones
  - Hypothesize this objective is suitable for abstractive summarization as it closely resembles the down-stream task, encouraging whole-document understanding and summary like generation
  - In practice, could apply both GSG and MLM simultaneously
  - But MLM does not improve down-stream tasks, so no included in PEGASUS-Large



제안하는 방법

- Selecting m gap sentences from a document
  - Random
    - Uniformly select m sentences at random
  - Lead
    - Select the first m sentences
  - Principal
    - Select top-m scored sentences according to importance
    - As a proxy for importance compute ROUGE1-F1 between sentence and rest of the document
    - (**IND**) $s_i = rouge(x_i, D \setminus \{x_i\}), \forall i$
    - (**SEQ**) Instead of scoring independently, it is also possible to greedily maximize ROUGE1-F1 between selected sentences and remaining sentences
    - Also, considering n-grams as a set(**UNIQ**), or double-counting identical n-grams(**Orig**) is both possible

Algorithm 1 Sequential Sentence Selection

$S := \emptyset$
for $j \leftarrow 1$ to m **do**
  $s_i := rouge\Big(S \cup \{x_i\}, D \setminus (S \cup \{x_i\})\Big)$
  $\forall i \;\; s.t. \;\; x_i \notin S$
  $k := \arg\max_i \{s_i\}_n$
  $S := S \cup \{x_k\}$
end for

제안하는 방법

- Selecting m gap sentences from a document

INVITATION ONLY We are very excited to be co-hosting a major drinks reception with our friends at Progress. This event will sell out, so make sure to register at the link above. Speakers include Rajesh Agrawal, the London Deputy Mayor for Business, Alison McGovern, the Chair of Progress, and Seema Malhotra MP. Huge thanks to the our friends at the ACCA, who have supported this event. The Labour Business Fringe at this year's Labour Annual Conference is being co-sponsored by Labour in the City and the Industry Forum. Speakers include John McDonnell, Shadow Chancellor, and Rebecca Long-Bailey, the Shadow Chief Secretary to the Treasury, and our own Chair, Kitty Ussher. Attendance is free, and refreshments will be provided.

Figure 2: An example of sentences (from the C4 corpus) selected by Random, Lead and Ind-Orig respectively. Best viewed in color.

실험 결과

- Pre-training Corpus

Pre-training Corpus	Size	Explanation
C4	750GB	Cleaned version of Common Crawl Consists of text from 350M Web-pages
HugeNews	3.8TB	Collected from news and news-like websites from 2013 ~ 2019 Only main article text was extracted as plain text

실험 결과

- Downstream Summarization Datasets

Downstream Summarization Datasets	Size	Explanation
Xsum	227k articles	BBC articles from 2010 ~ 2017 Covers wide variety of subject
CNN / DailyMail	93k articles (CNN) 220k articles (DM)	Use non-anonymized variant
NewsRoom	1.3M articles	Written by authors and editors in newsroom of 38 major publications between 1998 ~ 2017
Multi-News	56k articles	From site newser.com
Gigaword	4M articles	News articles from seven publishers
WikiHow	200k examples	Dataset of instructions from online WikiHow.com website

실험 결과

- Downstream Summarization Datasets

Downstream Summarization Datasets	Size	Explanation
arXiv, PubMed	113k examples (arXiv) 215k examples (PubMed)	Two long document of scientific publications from arXiv.org and PubMed
BigPatent	1.3M examples	U.S. patents under nine patent classification categories
Reddit TIFU	120k examples	Informal stories from online discussion forum Reddit TIFU sub-reddit from 2013-Jan to 2018-Mar
AESLC	18k examples	Collection of email messages of employees in Enron Corporation
BillSum	23k examples	US Congressional bills from 103^{rd}-115^{th} (1993 ~ 2018) sessions of Congress

실험 결과

- (Ablation Studies) Choosing best set of conditions for PEGASUS-Large
- Conduct Experiments with PEGASUS model on various datasets
- Zero and Low-Resource Summarization
- Qualitative Observations and Human Evaluation
- Test-set Overlap with Pre-training Corpus

실험 결과

- Evaluate choices of pretraining with PEGASUS-Base(223M)
  - Ablations Studies on most effective way of Pretraining
  - Pre-training corpus, Pre-training Objective, Vocabulary size
  - y-axis is normalized by using $\frac{1}{3}\left(\frac{R1}{R1_{base}}+\frac{R2}{R2_{base}}+\frac{R3}{R3_{base}}\right)$

- (Ablation Studies) Effect on Pre-training Corpus
  - HugeNews was more effective than C4 on two news downstream datasets
  - C4 was more effective than HugeNews on non-news informal datasets
  - Pretraniing models transfer more effectively to downstream tasks when domains align better

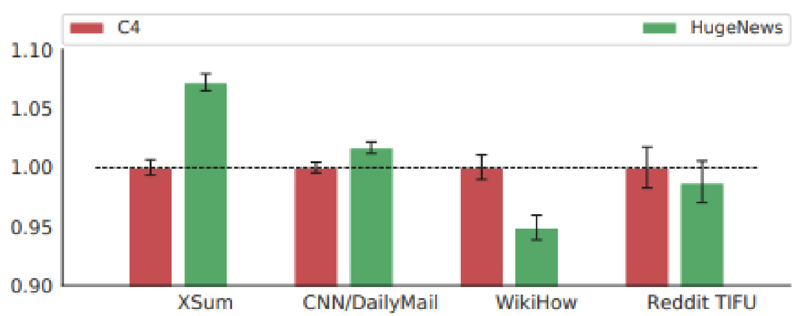
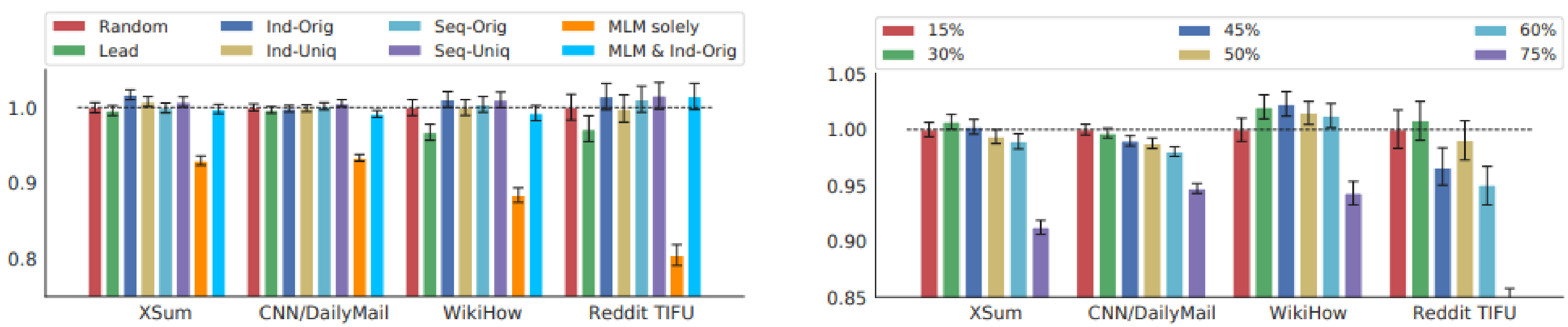


Figure 3: Effect of pre-training corpus. PEGASUS$_{BASE}$ pre-trained on C4 (350M Web-pages) and HugeNews (1.5B news-like documents).

실험 결과

- (Ablation Studies) Effect on Pre-training Objective
  - Compare six varients of GSG
  - Lead, Random, Ind-Orig, Ind-Uniq, Seq-Orig, Seq-Uniq
  - Ind-Orig achieved the best performance, and was chose for PEGASUS-Large
  - While MLM improved performance at early pre-training checkpoints, no more gain in further steps

  - Compare Gap-sentences Ratio(GSR) from 15% ~ 75%
  - Lower GSR makes pre-training less challenging and computationally efficient
  - Higher GSR loses contextual information necessary to guide generation
  - Different downstream datasets had slightly different optima
  - For PEGASUS-Large, 30% was chose



(a) Effect of pre-training objectives (30% GSR). (b) Effect of gap sentences ratio with GSG (Ind-Orig).

Figure 4: Effect of pre-training settings with PEGASUS$_{\text{BASE}}$ pre-trained on C4.

실험 결과

- (Ablation Studies) Effect on Tokenization Methods
  - Compare Byte-pair-encoding and SentencePiece Unigram algorithm
  - Evaluate Unigram with 32~256k
  - Models were pre-trained for 500k steps on C4 corpus with Ind-Orig objective and 15% GSR
  - BPE and Unigram were comparable on news datasets
  - Unigram outperformed BPE on non-news datasets
  - Unigram 96k was chose for PEGASUS-Large

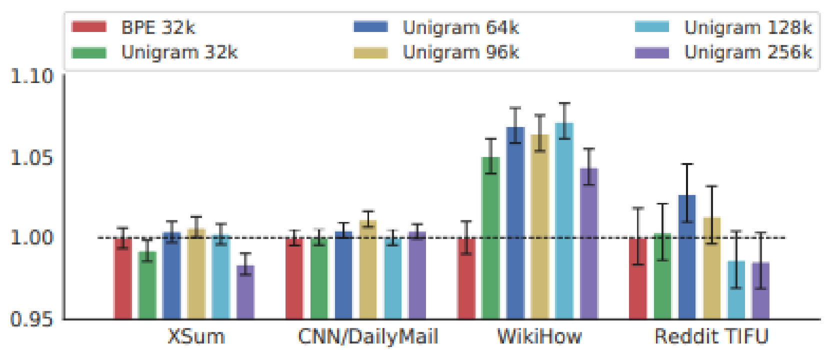


Figure 5: Effect of vocabulary with PEGASUS$_{\text{BASE}}$ trained on C4 (15% GSR, Ind-Orig).

실험 결과

- Conduct experiments on 4 of 12 datasets with PEGASUS-Large(568M)
  - XSum, CNN/DailyMail, WikiHow and Reddit TIFU
  - Adopted best practices found in PEGASUS-base ablation studies using GSG(Ind-Orig), 30% GSR, 96k sized Unigram Tokenizer
  - Supports both C4 version and HugeNews version which specializes on different tasks
  - On BigPatent, arXiv, PubMed and Multi-News, truncate into 256 tokens*
  - While PEGASUS-Base exceeded current SOTA, PEGASUS-Large achieved better than SOTA on all downstream datasets using HugeNews, except for on WikiHow

R1/R2/RL	Dataset size	$\text{Transformer}_{\text{BASE}}$	$\text{PEGASUS}_{\text{BASE}}$	Previous SOTA	$\text{PEGASUS}_{\text{LARGE}}$ (C4)	$\text{PEGASUS}_{\text{LARGE}}$ (HugeNews)
XSum	226k	30.83/10.83/24.41	39.79/16.58/31.70	45.14/22.27/37.25	45.20/22.06/36.99	**47.21/24.56/39.25**
CNN/DailyMail	311k	38.27/15.03/35.48	41.79/18.81/38.93	**44.16**/21.28/40.90	43.90/21.20/40.76	**44.17/21.47/41.11**
NEWSROOM	1212k	40.28/27.93/36.52	42.38/30.06/38.52	39.91/28.38/36.87	**45.07/33.39/41.28**	**45.15/33.51/41.33**
Multi-News	56k	34.36/5.42/15.75	42.24/13.27/21.44	43.47/14.89/17.41	46.74/17.95/24.26	**47.52/18.72/24.91**
Gigaword	3995k	35.70/16.75/32.83	36.91/17.66/34.08	**39.14/19.92/36.57**	38.75/**19.96**/36.14	**39.12/19.86**/36.24
WikiHow	168k	32.48/10.53/23.86	36.58/15.64/30.01	28.53/9.23/26.54	**43.06/19.71/34.80**	41.35/18.51/33.42
Reddit TIFU	42k	15.89/1.94/12.22	24.36/6.09/18.75	19.0/3.7/15.1	**26.54/8.94/21.64**	**26.63/9.01/21.60**
BIGPATENT	1341k	42.98/20.51/31.87	43.55/20.43/31.80	37.52/10.63/22.79	**53.63/33.16/42.25**	53.41/32.89/42.07
arXiv	215k	35.63/7.95/20.00	34.81/10.16/22.50	41.59/14.26/23.55	**44.70/17.27/25.80**	**44.67/17.18/25.73**
PubMed	133k	33.94/7.43/19.02	39.98/15.15/25.23	40.59/15.59/23.59	**45.49/19.90/27.69**	45.09/19.56/27.42
AESLC	18k	15.04/7.39/14.93	34.85/18.94/34.10	23.67/10.29/23.44	**37.69/21.85/36.84**	37.40/21.22/36.45
BillSum	24k	44.05/21.30/30.98	51.42/29.68/37.78	40.80/23.83/33.73	**57.20**/39.56/**45.80**	**57.31/40.19/45.82**

실험 결과

- Zero and Low-Resource Summarization
  - In real world, it is difficult to collect a large number of supervised examples to train or fine-tune
  - To simulate low resource summarization, pick 10^k(k=1,2,3,4) training examples from dataset
  - PEGASUS achieves remarkable performance with small number of fine-tuning data

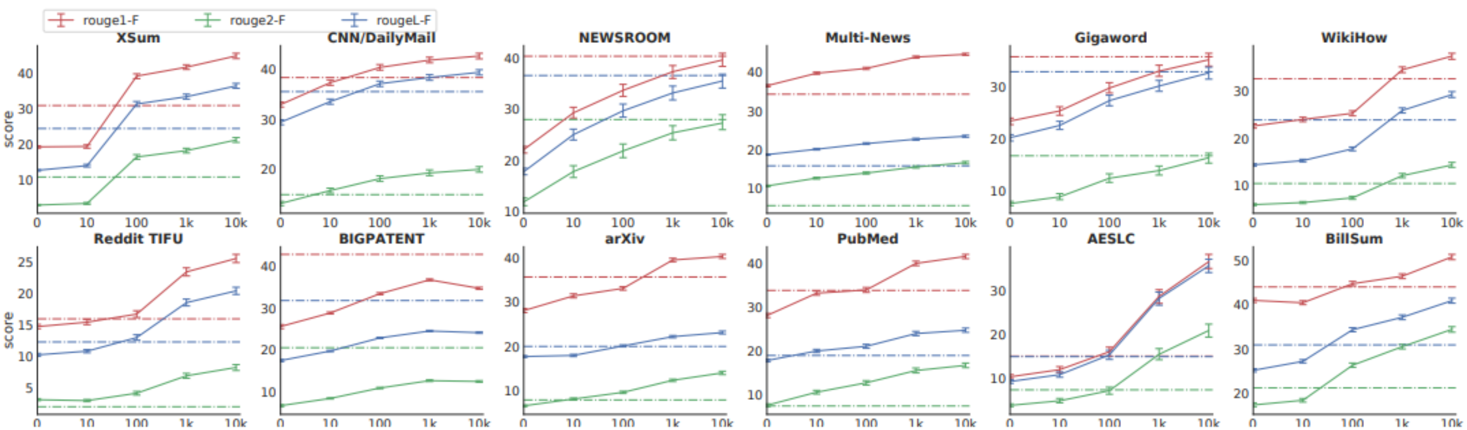


Figure 6: Fine-tuning with limited supervised examples. The solid lines are $\text{PEGASUS}_{\text{LARGE}}$ fine-tuned on 0 (zero shot), 10, 100, 1k,10k examples. The dashed lines are $\text{Transformer}_{\text{BASE}}$ models, equivalent in capacity as $\text{PEGASUS}_{\text{BASE}}$ and trained using the full supervised datasets, but with no pre-training. All numbers are reported in Appendix E.

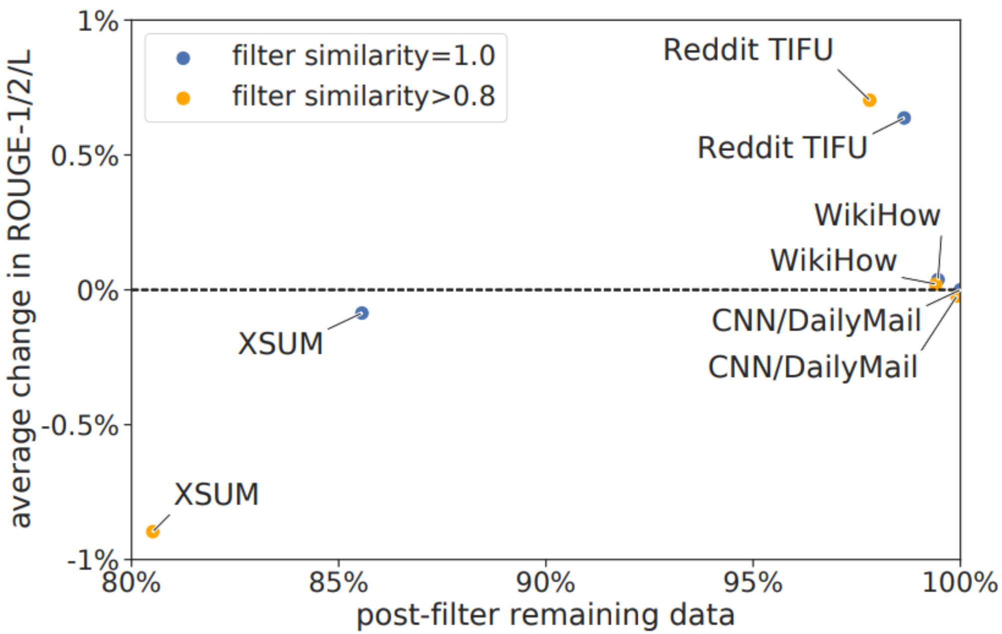
실험 결과

- Qualitative Observations and Human Evaluation
  - Repetitive text in model outputs were not found
  - ROUGE clearly has draw-backs
    - *Kryściński, Wojciech, et al. "Neural Text Summarization: A Critical Evaluation." Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019.*
  - Choosing perplexity-optimized models using aggregated ROUGE(rather than directly optimizing ROUGE) resulted in qualitatively good models
  - Using Amazon Mechanical Turk, compared model summaries with human reference summaries*

Datasets	**XSum** mean (p-value)	**CNN/DailyMail** mean (p-value)	**Reddit TIFU** mean (p-value)
Experiment 1: pretrain comparison			
Human-written	3.0 (-)	3.1 (-)	3.2 (-)
$\text{PEGASUS}_{\text{LARGE}}$ (HugeNews)	**3.0** (0.6)	**3.6** (0.0001)	**3.2** (0.7)
$\text{PEGASUS}_{\text{LARGE}}$ (C4)	**3.1** (0.7)	**3.5** (0.009)	**3.1** (0.3)
$\text{Transformer}_{\text{BASE}}$	2.0 (3e-10)	**2.9** (0.06)	1.4 (5e-23)
Experiment 2: low resource			
Human-written	3.2 (-)	3.2(-)	3.3 (-)
$\text{PEGASUS}_{\text{LARGE}}$ (HugeNews) 10 examples	**2.8** (0.1)	**3.4** (0.007)	2.6 (0.006)
$\text{PEGASUS}_{\text{LARGE}}$ (HugeNews) 100 examples	**3.2** (0.5)	**3.4** (0.08)	2.1 (4e-8)
$\text{PEGASUS}_{\text{LARGE}}$ (HugeNews) 1000 examples	**3.4** (0.3)	**3.6** (0.07)	2.7 (0.01)
$\text{PEGASUS}_{\text{LARGE}}$ (HugeNews) full supervision	**3.4** (0.3)	**3.3** (0.1)	**2.8** (0.05)

실험 결과

- Test-set Overlap with Pre-training Corpus
  - Measuring the extent of overlap between the pre-training corpus and downstream datasets
  - Use ROUGE-2 recall as a similarity measure
  - Only XSum has significant amount of overlap 15% to 20%, and filtering those examples does not change ROUGE score more than 1%
  - There was no clear memorization



실험 결과

- Screenshot of Amazon Mturk HIIT
  - Evaluate 3 models (PEGASUS-Large on C4, PEGASUS-Large on HugeNews, Transformers-base) and human summary

Read the document below, then rate the summaries for quality on a scale of 1-5. (1 = Poor summary, 5 = Great summary)

Document:
Tynan, a former Manchester City player, died after being hit by a train at West Allerton station in Merseyside on Tuesday, British Transport Police said. Tynan's death is not being treated as suspicious. Her family paid tribute to a "vibrant, generous and fun-loving girl", who was "a dedicated athlete, never happier than when she had a ball at her feet". Tynan began her career at Liverpool Feds, spent six years at Everton's Centre of Excellence and was playing for Women's Premier League side Fylde Ladies. A family statement also said she was a "the most loving and caring daughter and sister anyone could wish for" and that she was the "ultimate team player". It added: "Zoe always knew how to cheer anyone up, and was a loyal, straight-talking friend to many. She touched so many people's lives and will never be forgotten." Tynan joined Manchester City in 2015, making one Women's FA Cup appearance before moving to Fylde. Floral tributes have been left at the scene, according to the Liverpool Echo. England internationals including Lucy Bronze and Casey Stoney have also paid tribute. Fylde manager Luke Swindlehurst said: "We want to remember Zoe in the best possible way: a hugely talented player and an immensely likable character." Tynan had appeared for England at various youth levels and was recently included in the Under-19 squad for a training camp at St George's Park. The Football Association said it was "deeply saddened" by the death and Tynan's Under-19 coach Mo Marley described her as a "hugely-liked and popular member of the team".

Summary:
England Under-19 Women's and Fylde Ladies midfielder Zoe Tynan has died, aged 18.

Summary:
England Under-19 midfielder Zoe Tynan has been struck and killed by a train.

Summary:
England Under-19 midfielder Zoe Tynan has died after being struck by a train.

Summary:
A 27-year-old woman has been mugged in Liverpool by two men who stole her wallet. A family statement also said she was a " the most loving and caring daughter and sister anyone could wish for" and that she was the "ultimate team player".

Figure F.1: A screenshot of the Amazon MTurk HIIT.

실험 결과

- Example of summary with low ROUGE-2, but qualitatively good

Document: chelsea will face paris saint-germain, the french team who knocked jose mourinhos side out of the champions league this season, in a pre-season friendly in july. the blues, who were sent crashing out on away goals at the last-16 stage following a 2-2 draw at stamford bridge, will play psg in north carolina on july 25. it is one of three games mourinhos side will feature in across the pond as they gear up to defend a probable premier league title. john terry leads the celebrations as chelsea close in on the premier league title with a 0-0 draw at arsenal . eden hazard, the pfa player of the year, will line-up for chelsea when they travel to the usa in the summer . new york red bulls - july 22 - new jersey . paris saint-germain - july 25 - charlotte, north carolina . barcelona - july 28 - washington d.c. fiorentina - august 5 - stamford bridge . chelsea, 10 points ahead of arsenal with just four games to play, will also face the new york red bulls on july 22 and spanish giants barcelona six days later in washington. chelsea fans will then get to see their side before the premier league campaign kicks-off with a friendly against fiorentina at stamford bridge on august 5. all four matches mark chelseas participation in this summers pre-season international champions cup with manchester united, who mourinhos side will not face, la galaxy, porto and san jose earthquakes also involved. im pleased we are able to announce our fixtures for what promises to be an exciting summer,' said chelsea chairman bruce buck. as promised, we face some excellent opposition across several iconic venues in the united states and to top it off we are delighted to be hosting fiorentina at stamford ... ...

Ground-truth: chelsea to play three matches inside six days in the united states . they will face new york red bulls, paris saint-germain and barcelona . fiorentina will then travel to stamford bridge for friendly on august 5 . four matches will make up chelsea's participation in champions cup . read: chelsea interested in 43m antoine griezmann .

Model: jose mourinho's side will play psg in north carolina on july 25 . chelsea will also face the new york red bulls and barcelona . the blues will play fiorentina at stamford bridge on august 5 .

Figure G.1: A CNN/DailyMail PEGASUS$_{\text{LARGE}}$ model summary with relatively low ROUGE2-F of 16, but qualitatively quite good, and factually accurate.

실험 결과

- Example of summaries generated by PEGASUS

Document (ID #289)	machine learning methods are used widely within high energy physics (hep) . one promising approach , used extensively outside of hep for applications such as handwriting recognition , is that of support vector machines (svms) , a supervised learning model used with associated learning algorithms for multivariate analysis (mva) . developed originally in the 1960s , with the current standard version proposed in 1995 xcite , svms aim to classify data points using a maximal margin hyperplane mapped from a linear classification problem to a possibly infinite dimensional hyperspace . however this means svms , like other mva classifiers , have a number of free parameters which need to be tuned on a case by case basis . this motivates a number methods for ensuring the classifier is sufficiently generalised such that when used on an unseen dataset the performance can be accurately predicted . in this paper a brief overview of svms is given in section [sec : svm] , with an example using svms shown in section [sec : checker] . generalisation is discussed in section [sec : gen] with an illustrative example of how this can improve performance given in section [sec : check2] . consider the problem of linear classification with the svm where the training set , xmath0 , is linearly separable . we define a separating hyperplane given by xmath1 , where xmath2 , the weight vector , is perpendicular to the hyperplane , and xmath3 , the bias , determines the distance of the hyperplane from the origin (fig .) . a decision function defined by xmath4 is used to positively and negatively classify xmath5 , the points in the training set . without further constraint the ... ...
Gold	we review the concept of support vector machines (svms) and discuss examples of their use . one of the benefits of svm algorithms , compared with neural networks and decision trees is that they can be less susceptible to over fitting than those other algorithms are to over training . this issue is related to the generalisation of a multivariate algorithm (mva) ; a problem that has often been overlooked in particle physics . we discuss cross validation and how this can be used to improve the generalisation of a mva in the context of high energy physics analyses . the examples presented use the toolkit for multivariate analysis (tmva) based on root and describe our improvements to the svm functionality and new tools introduced for cross validation within this framework .
Model	support vector machines (svms) are one of the most promising methods for machine learning in high energy physics . svms aim to classify data points using a maximal margin hyperplane mapped from a linear classification problem to a possibly infinite dimensional hyperspace . however this means svms , like other mva classifiers , have a number of free parameters which need to be tuned on a case by case basis . this motivates a number methods for ensuring the classifier is sufficiently generalised such that when used on an unseen dataset the performance can be accurately predicted . in this paper a brief overview of svms is given , with an example using svms shown in section [sec : checker] . generalisation is discussed with an illustrative example of how this can improve performance given in section [sec : gen] .
ROUGE1-F1	35.29

결론

- Instead of truncation, scaling up L_{input} or applying two-stage approach may improve performance even more
  - *Liu, Peter J., et al. "Generating Wikipedia by Summarizing Long Sequences." International Conference on Learning Representations. 2018.*

- Gap-Sentence Selection methods is a SOTA pretraining method for abstractive summarization task

- PEGASUS was able to adapt to unseen summarization datasets very quickly

- Model summaries achieved human performance on multiple datasets using human evaluation

Appendix : 과제 진행 과정

Inference using Existing Pre-trained Models

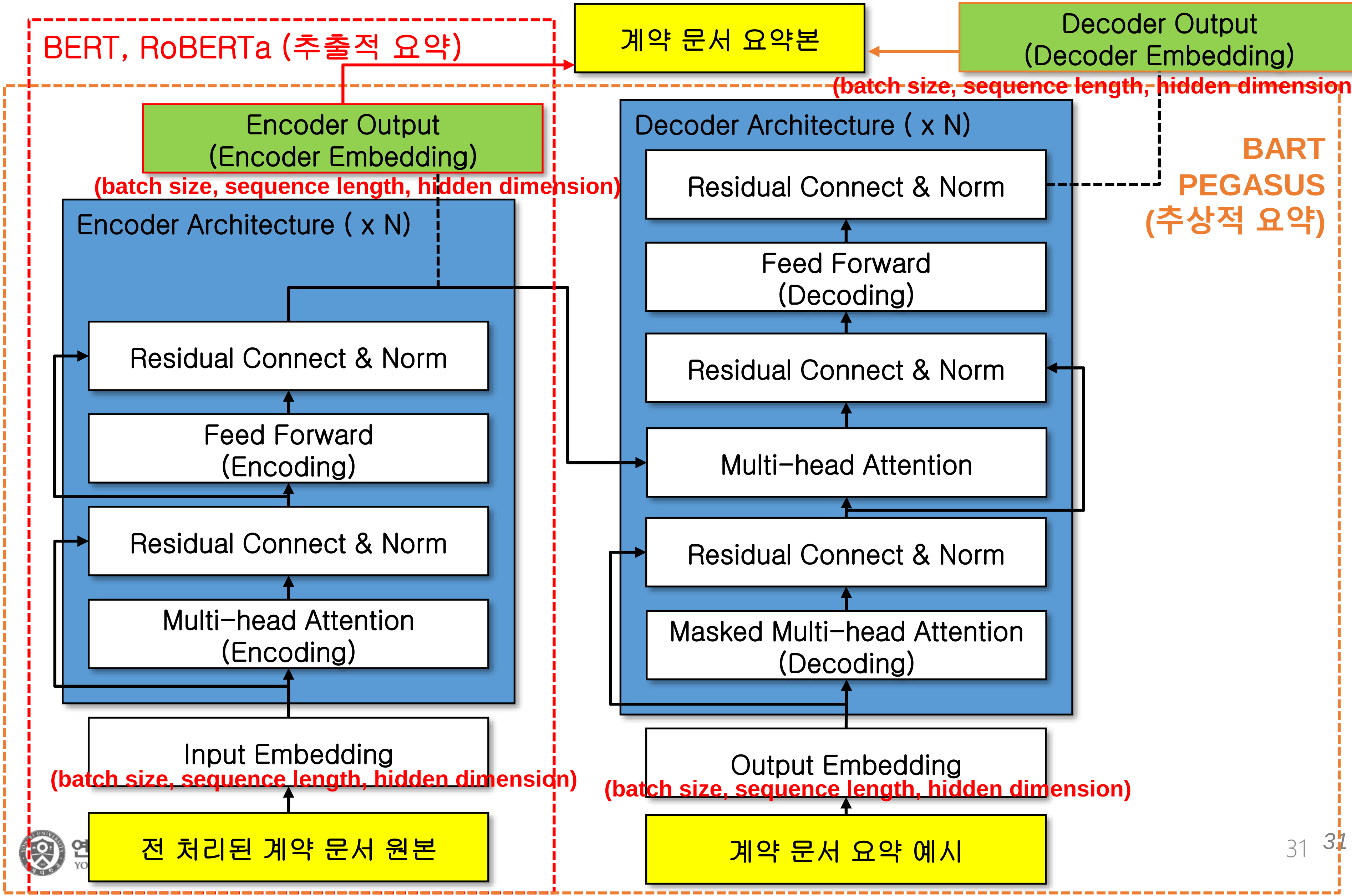
Fine-tuning Using Existing Pre-trained Models

Pre-training new model from scratch

Hugging Face

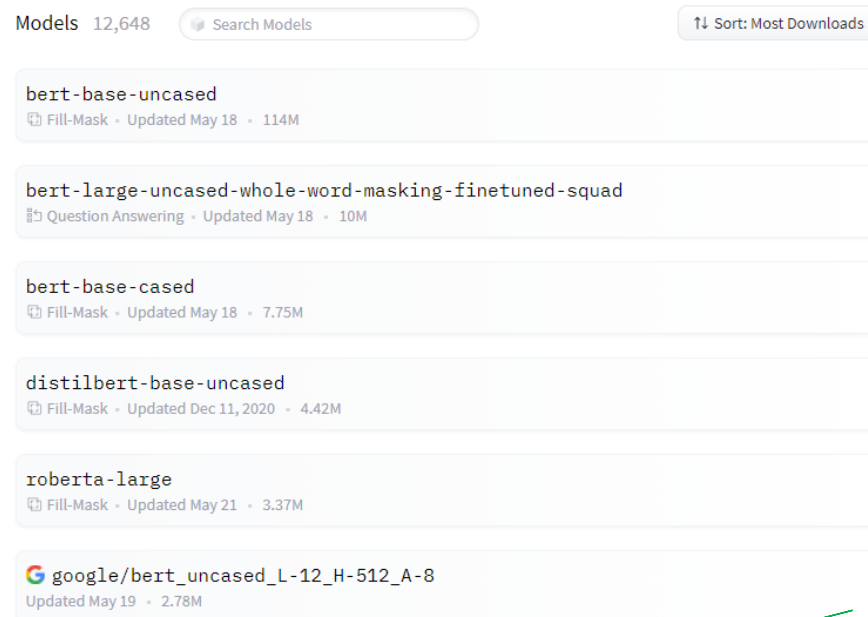
PyTorch

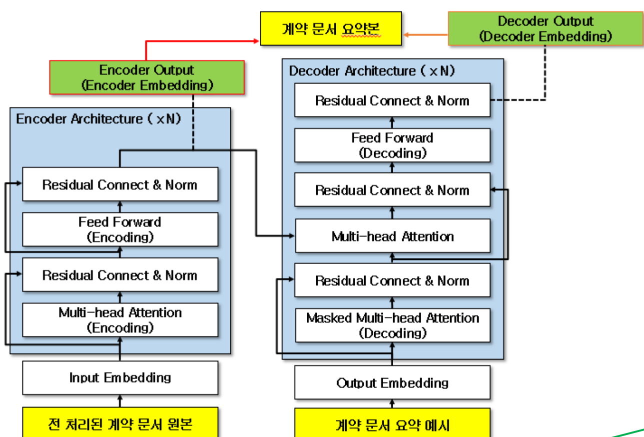
Appendix : 과제 진행 과정



Appendix : 과제 진행 과정

- Inference with Existing Models
    - Load model from HuggingFace Library



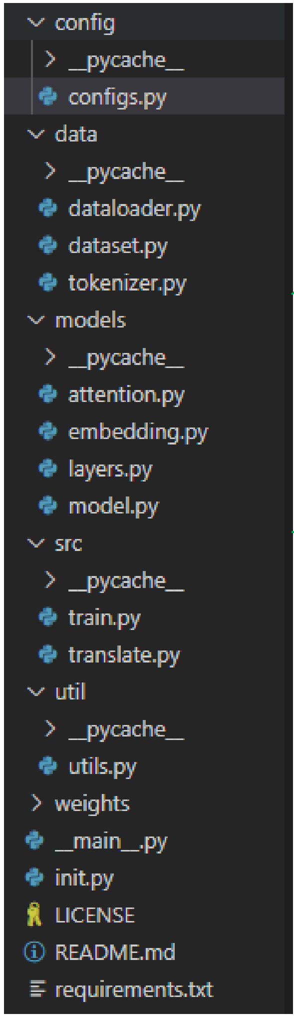


Hugging Face

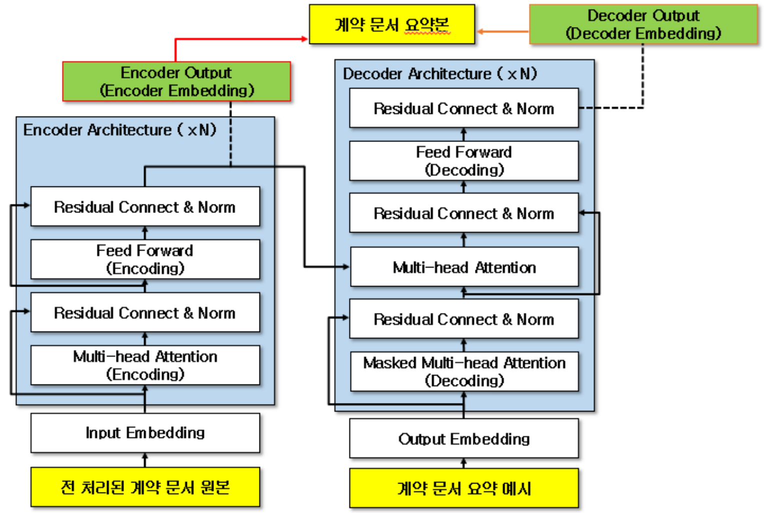
```python
from transformers import AutoTokenizer, AutoModelForMaskedLM

tokenizer = AutoTokenizer.from_pretrained("roberta-base")

model = AutoModelForMaskedLM.from_pretrained("roberta-base")
```

Appendix : 과제 진행 과정

- Fine-tuning existing Pre-trained Models



PyTorch



Trainer

DataLoader

Tokenizer

Loss function
Metric

Appendix : 과제 진행 과정

- 원문 예시

```
text2 = """
We may cancel, suspend, or terminate your Account and your access to your Trading Items, Virtual
Money, Virtual Goods, the Content, or the Services, in our sole discretion and without prior notice,
including if (a) your Account is inactive (i.e., not used or logged into) for one year; (b) you fail to
comply with these Terms; (c ) we suspect fraud or misuse by you of Trading Items, Virtual Money,
Virtual Goods, or other Content; (d) we suspect any other unlawful activity associated with your
Account; or (e) we are acting to protect the Services, our systems, the App, any of our users, or
the reputation of Niantic, TPC, or TPCI. We have no obligation or responsibility to, and will not
reimburse or refund, you for any Trading Items, Virtual Money, or Virtual Goods lost due to such
cancellation, suspension, or termination. You acknowledge that Niantic is not required to provide a
refund for any reason, and that you will not receive money or other compensation for unused Virtual
Money and Virtual Goods when your Account is closed, whether such closure was voluntary or
involuntary. We have the right to offer, modify, eliminate, and/or terminate Trading Items, Virtual
Money, Virtual Goods, the Content, and/or the Services, or any portion thereof, at any time, without
notice or liability to you. If we discontinue the use of Virtual Money or Virtual Goods, we will
provide at least 60 days advance notice to you by posting a notice on the Site or App or through other
communications.
"""
```

Appendix : 과제 진행 과정

- 요약 예시

- BART

```
We may cancel, suspend, or terminate your Account and your access to your Trading Items, Virtual worrisomeMoney, Virtual Goods, t
he Content, or the Services, in our sole discretion and without prior notice. We have no obligation or responsibility to, and wil
l notreimburse or refund, you for any Trading items, Virtual Money, or Virtual Goods lost due to such cancellation, suspension, o
r termination.
```

- T5

```
we may cancel, suspend, or terminate your Account in our sole discretion . we have no obligation or responsibility to, and will n
ot reimburse or refund, you . you acknowledge that Niantic is not required to provide a refund .
```

- Pegasus

```
We may cancel, suspend, or terminate your Account and your access to your Trading Items, Virtual Money, Virtual Goods, the Conten
t, or the Services, in our sole discretion and without prior notice .<n>We have the right to offer, modify, eliminate, and/or ter
minate Trading Items, Virtual Money, Virtual Goods, the Content, and/or the Services, or any portion thereof, without notice or l
iability to you .
```

Appendix : 과제 진행 과정

- 성능 측정표 예시

- Rogue 점수

Benchmark Dataset	CNN / DailyMail			Xsum		
Metrics	RG-1	RG-2	RG-L	RG-1	RG-2	RG-L
BART – Large	44.16	21.28	40.90	45.14	22.27	37.25
T5 – Large	43.52	21.55	40.69	-	-	-
PEGASUS - Large	44.17	21.47	41.11	47.21	24.56	39.25