질의 기반 의미그래프와 적응형 메타학습을 이용한 생성형 요약 시스템

고 영 중

Youngjoong Ko

Computer Science & Engineering, Graduate School of AI

SKKU NLP LAB

UNIVERSITY

Contents

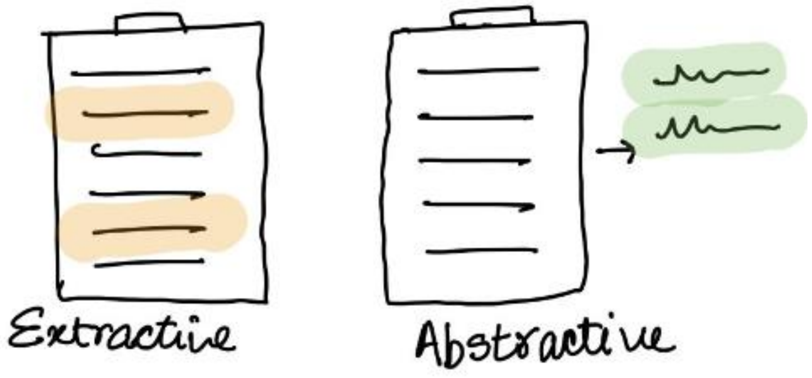
Text Summarization

❖ **What's Text Summarization (TS)**

- Reducing the size of a document while preserving its content
- to extract content and present the most important content to a user in a condensed form

❖ **Text Summarization System**

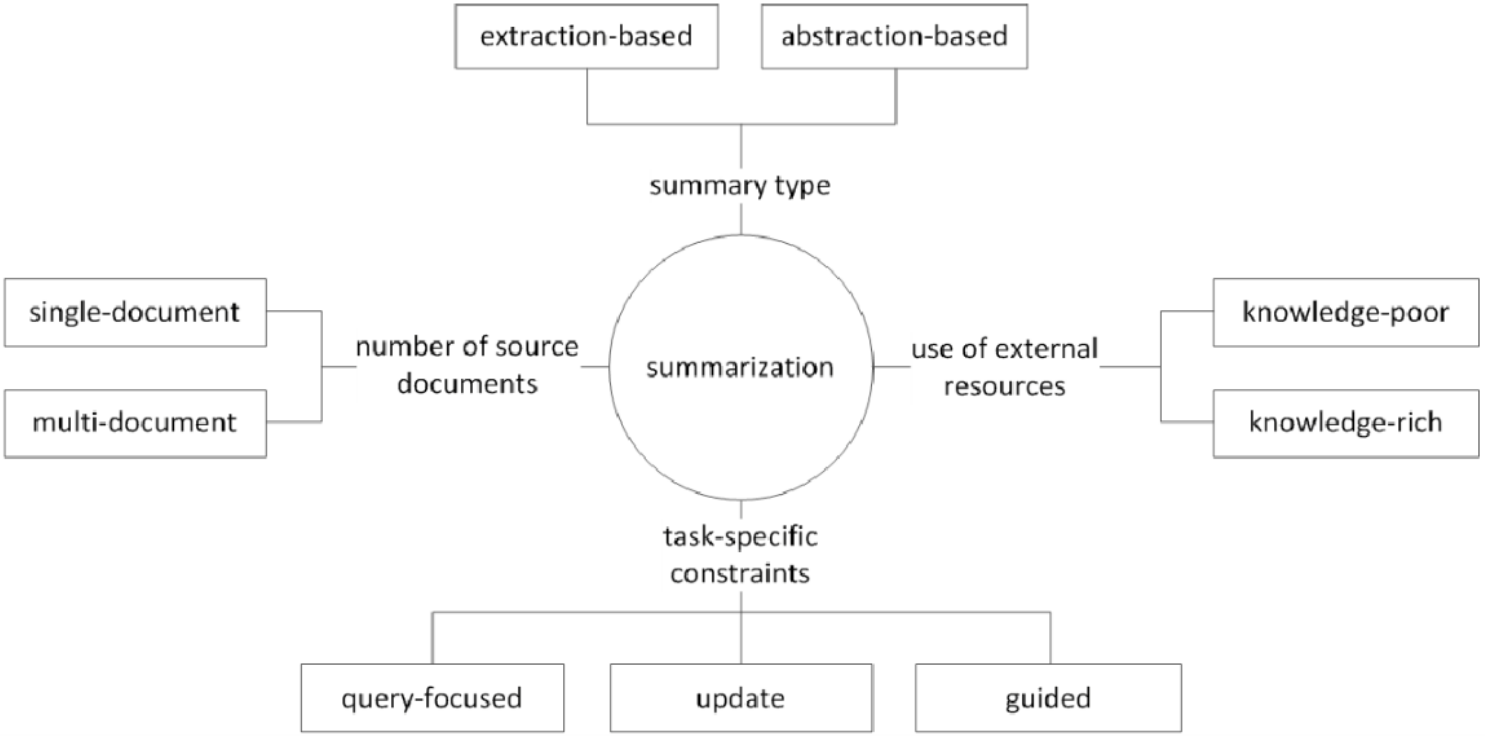
- Identify the most salient information in a document
- Most widespread summarization strategy
  - Extractive -> Abstractive Summarization



https://supkoon.tistory.com/40

Text Summarization

❖ Summarization Tasks



G. Sizov, "Extraction-Based Automatic Summarization: Theoretical and Empirical Investigation of Summarization Techniques"

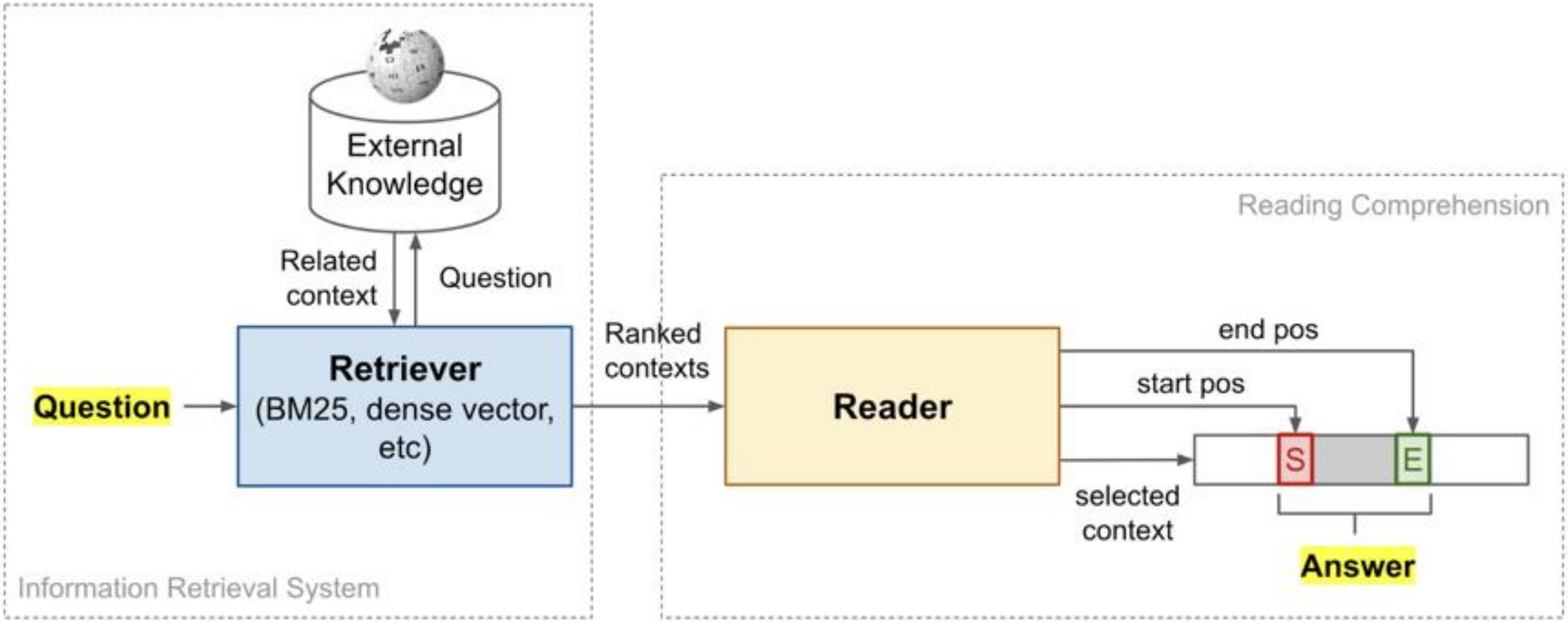
Text Summarization

❖ **Main Topics**

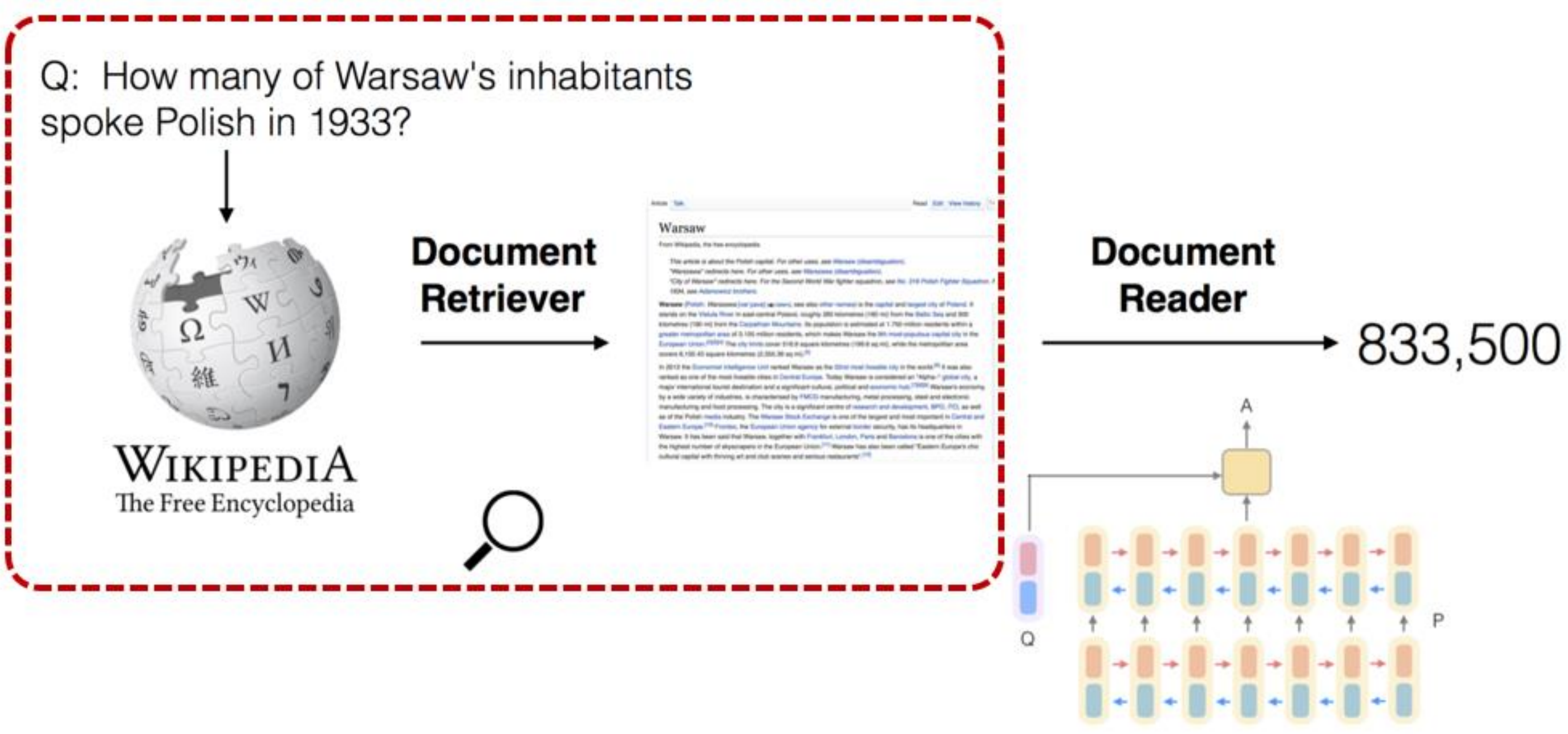
- Multi documents summarization (MDS)/ long document summarization
- **Text summarization with pretraining model**
- **Knowledge-enhanced text generation**
- **Few-shot learning (Meta Learning)**
- **Query-focused summarization (QFS)**
- Aspect-based summarization
- Update summarization
- Conversation summarization

Text Summarization for XAI

❖ **Open Domain QA**



❖ **Open Domain QA**



❖ 정답 추론 근거 생성

- 질의응답 시스템의 정답 추론 근거 생성
- 사용자에게 텍스트 형태의 예측 이유/근거 제공

예시)

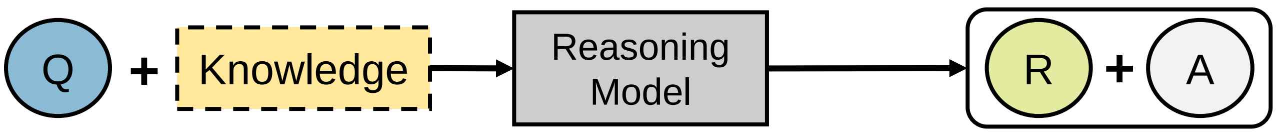
StrategyQA:

Question	Did Aristotle use a laptop?
Answer	No.
Rationale	Aristotle was alive until … while the first laptop was invented in …

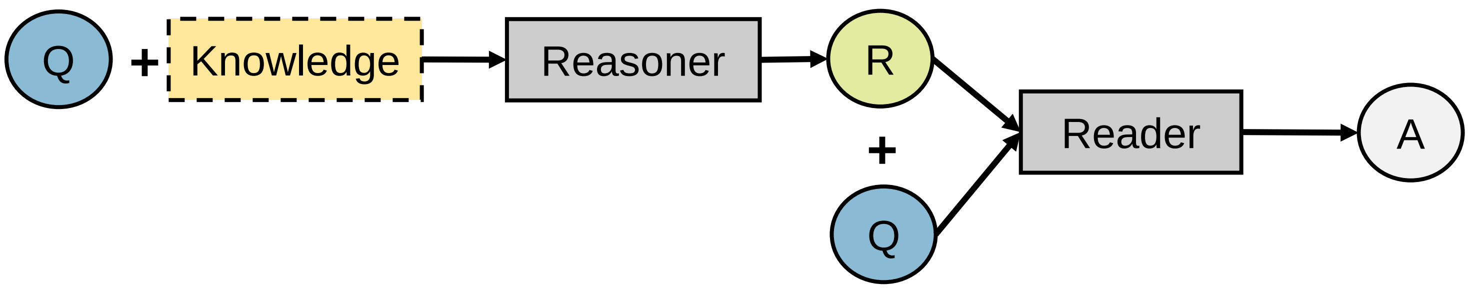
❖ 정답 추론 근거 생성

- **시스템 구조**

1. Self-rationalizing paradigm



2. Pipeline-rationalizing paradigm



- Question (Q): 질문 / Rationale (R): 근거 / Answer (A): 정답
- Knowledge: 텍스트 코퍼스, 대형 사전학습 언어모델, 지식 그래프…

Focused points for my research

❖ **Query Focused Text Summarization**

- 긴 문서에서 질의와 관련 있는 중요한 내용을 함축하여 요약 생성

(Query)
Can corporal punishment cause physical damage?

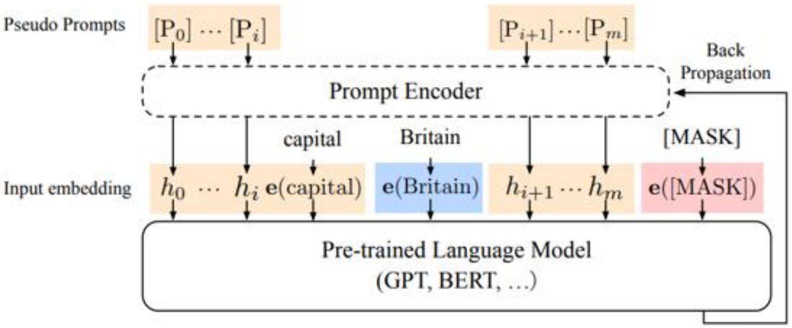
(Document)
The actual physical damage inflicted via corporal punishment on children can be horrifying. Exampl es can be found of students needing treatment for broken arms nerve and muscle damage and cerebr al hemorrhage. Spanking of the buttocks can cause damage to the sciatic nerve and therefore the leg to which it leads.

(Summary)
Corporal punishment can cause serious physical damage.

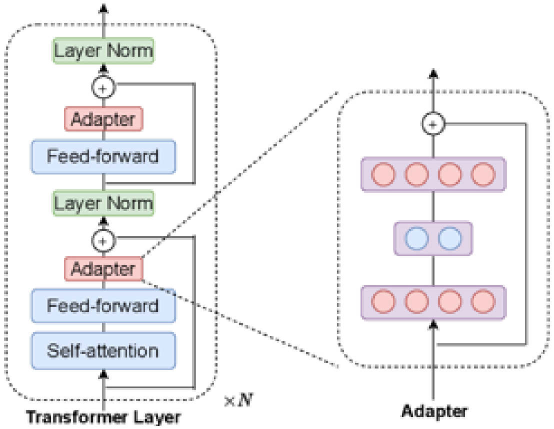
Focused points for my research

❖ Parameter Efficient Tuning

- In-context learning (P-tuning)
  - Liu., X. et al., GPT Understands, Too (P-tuning), arXiv, 2021.



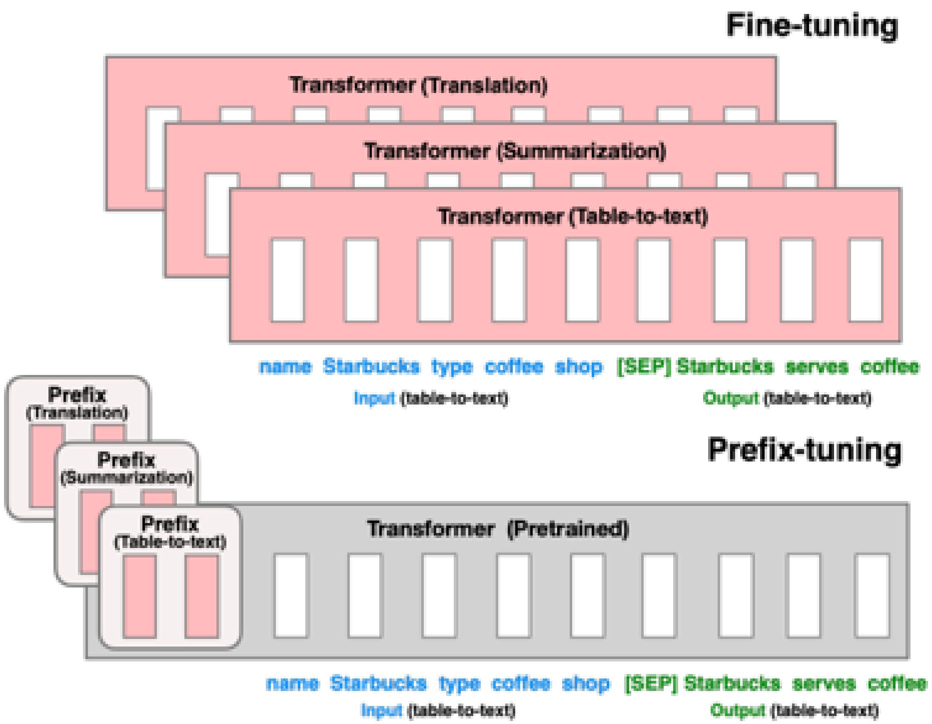
- Adapter-tuning
  - He, R., On the Effectiveness of Adapter-based Tuning for Pretrained Language Model Adaptation, ACL 2021.



Focused points for my research

❖ **Parameter Efficient Tuning**

- **Prefix Tuning**
  - PXiang L. L. et al., refix-tuning: Optimizing Continuous Prompts for Generation, ACL 2021.



Query-Focused Summarization

❖ Query-Focused Summarization

- Task to generate a summary, which can answer a query from the essential information of a source document.
- Query-focused summarization has been in spotlight because of the need to generate a summary suitable for a specific topic or user's interest.

(Query) Is solar energy generally environmentally friendly?
(Document) Because sun's rays are diffuse, solar panels must occupy substantial territory to generate any significant quantities of power. As a result, solar energy is land-intensive and creates a pressure to clear land of trees and vegetation to make way for solar panels. Owners of solar panels on home rooftops may also have an incentive to cut-down trees that are blocking solar panels from the sun's rays. This is a significant ecological threat.
(Gold Summary) Land-intensive solar power incentivizes clearing land ecosystems.
(BART Summary) Solar energy is land-intensive.
(QSG BART Summary) Solar energy is land-intensive which causes deforestation.

Natural Language Processing

Query-Focused Summarization

❖ Limitations of previous works

- Simple LM has a difficulty in directly reflecting relationships between important distant words just on the self-attention mechanism.
- Simple LM does not have any function to properly focus on the query information of QFS.

❖ QSG Transformer

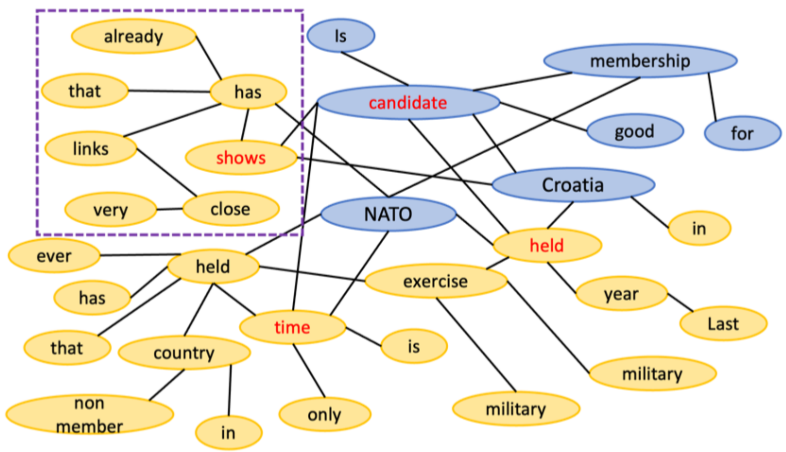
- QSG Transformer is a novel QFS model with Query-attentive Semantic Graph (QSG).
- It effectively utilizes relation information between distant words and query information using QSG.

Methodology

❖ Query-attentive Semantic Graph

Query: Is Croatia a good candidate for NATO membership?
Document: Last year NATO held a military exercise in Croatia. This is the o nly time that NATO has ever held a military exercise in a non member countr y. This shows that NATO already has very close links.
Ground Truth Summary: NATO already has close links with Croatia.

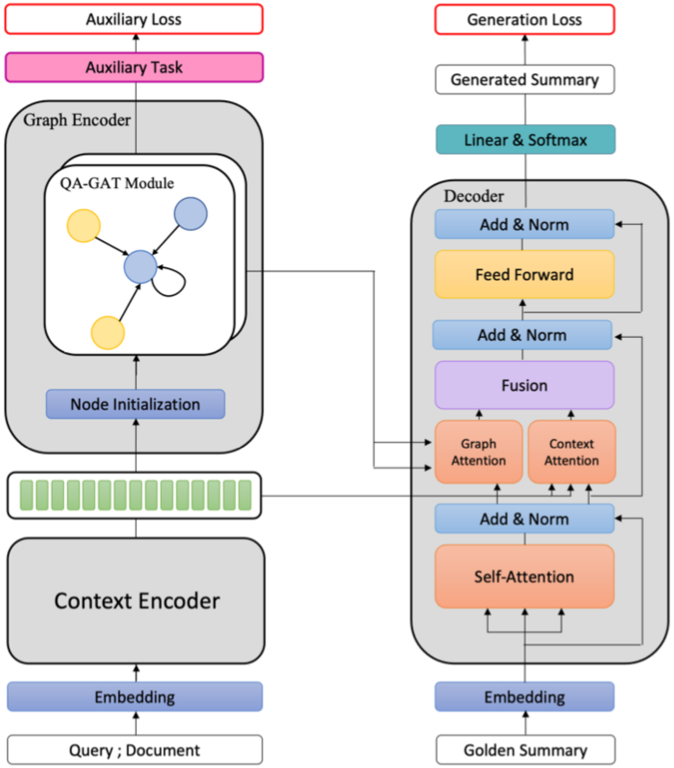
Query-attentive Semantic Graph



- POS tag, dependency tree and coreference resolution chain information are used to construct QSG.
- A root word of the query and the root words of sentences are connected.

Methodology

❖ QSG Transformer



- To utilize the information of QSG, QSG Transformer has 2 encoders and decoder generates a summary using context representations and graph node representations.

Methodology

❖ Graph Encoder

- **Query-Attentive GAT Module**

  - Query-attentive GAT module consists of a query-attentive PPR layer and the stack of several query-attentive GAT layers.
  - The importance score of each node for query nodes can be measured using query-attentive Personalized PageRank algorithm in the query-attentive PPR layer.
  - In query-attentive GAT layer, each node representation is updated using the conventional attention weight as well as the importance score in query-attentive PPR layer.

- **Auxiliary Task**

  - To strengthen the quality of the node representation, auxiliary self-supervised node classification task is added.
  - Each node is labeled by calculating shortest path length from each node to the query node.

Experiments & Results

❖ Main Results

	Debatepedia			PubMedQA		
Model	R-1	R-2	R-L	R-1	R-2	R-L
Transformer	41.7	33.6	41.3	30.4	8.4	22.3
CSA Transformer	46.4	**37.5**	45.9	-	-	-
QSG Transformer	**49.6**	37.0	**48.2**	**33.6**	**11.8**	**25.5**
SD2	41.3	18.8	40.4	32.3	10.5	26.0
QR-BERTSUM-TL	48.0	45.2	57.1	-	-	-
MSG	-	-	-	37.2	14.8	**30.2**
BART	58.0	43.6	56.3	38.2	16.0	28.9
BART-QFS	59.0	44.6	57.4	-	-	-
QSG BART	**64.9**	**52.3**	**63.3**	**38.4**	**17.0**	29.8

❖ Analysis

	R-1	R-2	R-L	R-1	R-2	R-L
Method	w/o auxiliary task			w/ auxiliary task		
GCN	46.2	34.7	45.2	46.9	34.6	45.7
GAT	45.7	32.9	44.3	48.0	35.3	46.6
PPNP	47.9	36.0	46.5	47.6	35.7	46.4
query-attentive GAT	48.3	36.2	47.0	**49.6**	**37.0**	**48.2**

Low-Resource Abstractive Summarization

❖ Need for low-resource abstractive summarization

- Most summary datasets consist of a specific domain like news.
- There can be few labeled documents in most cases of new domains.
- Annotating document-summary pairs is too expensive.

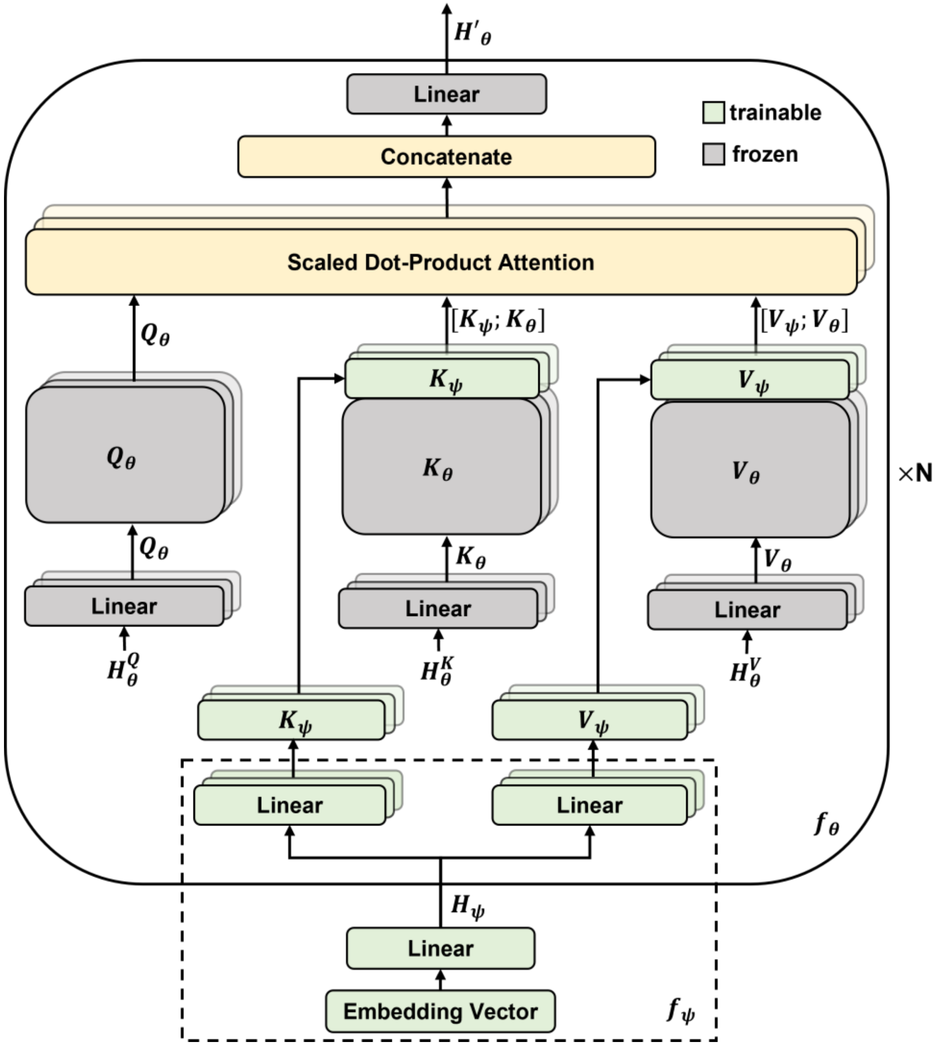
❖ Lightweight Meta-Learning for Low-Resource Abstractive Summarization

- Model-Agnostic Meta-Learning(MAML) is an algorithm to learn a new domain with a small number of examples.
- The large pre-trained model cannot be adapted well to small data. To address this, we introduce a lightweight module.

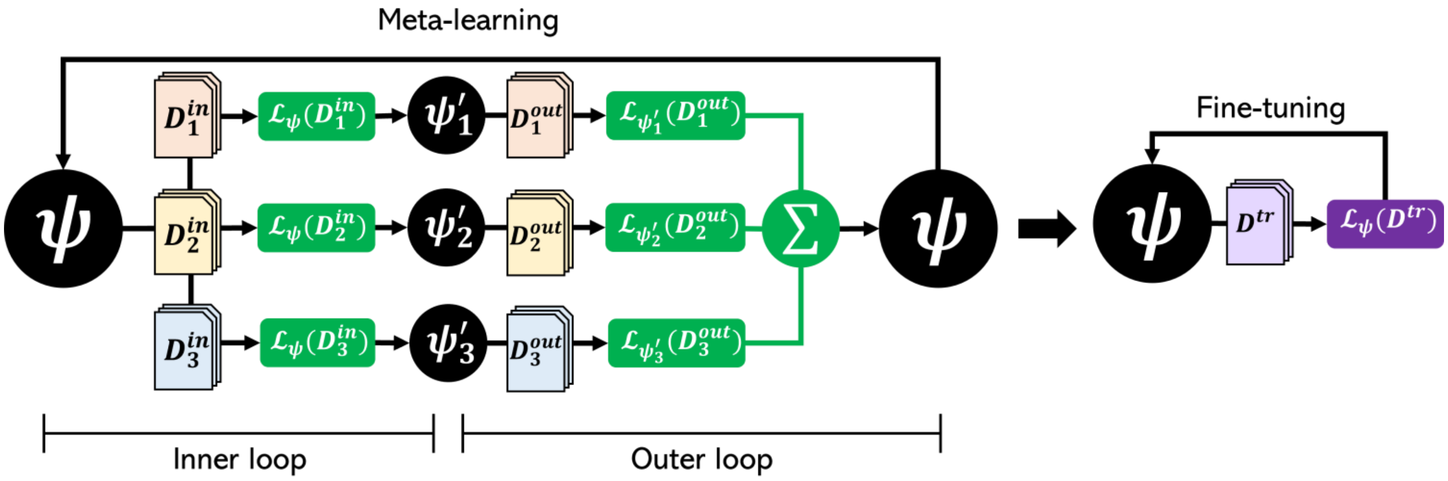
Methodology

❖ Lightweight Module

- Overview of the attention mechanism with the lightweight module.
- The dotted line box is the lightweight module.
- During meta-learning and fine-tuning, only the module is trainable.



❖ Training Framework



- In the inner loop, the module learns domain-specific knowledge.
- In the outer loop, the parameters of the module are initialized to fast adapt to the target domain using losses of domain adapted modules.
- While repeating the loops, the parameters of the module are initialized to quickly adapt to the target domain.

Experiments & Results

❖ Experiments

- Metric: ROUGE score(R-1 / R-2 / R-L)

Dataset	Labeled examples	PEGASUS R-1 / R-2 / R-L	MTL-ABS R-1 / R-2 / R-L	**Ours** R-1 / R-2 / R-L	Improving ratio R-1 / R-2 / R-L
AESLC	10	11.97/4.91/10.84	21.27/10.79/20.85	**24.25/11.62/23.49**	+14% / +8% / +13%
	100	16.05/7.20/15.32	23.88/12.06/23.18	**30.00/15.14/29.11**	+26% / +26% / +26%
BillSum	10	40.48/18.49/27.27	41.22/18.61/26.33	**46.64/25.07/30.90**	+13% / +35% / +13%
	100	44.78/26.40/**34.40**	45.29/22.74/29.56	**48.18/27.18**/33.28	+6% / +3% / -3%
Gigaword	10	25.32/8.88/22.55	28.98/11.86/26.74	**30.99/12.81/28.41**	+7% / +8% / +6%
	100	29.71/12.44/27.30	30.03/12.70/27.71	**31.54/13.61/29.03**	+5% / +7% / +5%
Multi-News	10	39.79/12.56/20.06	38.88/12.78/19.88	**43.60/14.85/20.70**	+10% / +16% / +3%
	100	41.04/13.88/21.52	39.64/13.64/20.45	**45.05/15.99/21.90**	+10% / +15% / +2%
Reddit-TIFU	10	15.36/2.91/10.76	18.03/**6.41**/17.10	**22.91**/6.06/**17.50**	+27% / -5% / +2%
	100	16.64/4.09/12.92	20.14/**7.71**/19.38	**25.37**/7.05/**19.81**	+26% / -9% / +2%
arXiv	10	31.38/8.16/17.97	35.81/10.26/20.51	**41.09/13.84/22.10**	+15% / +35% / +8%
	100	33.06/9.66/20.11	37.58/10.90/20.23	**41.30/13.97/22.54**	+10% / +28% / +11%
PubMed	10	33.31/10.58/20.05	34.08/10.05/18.66	**38.32/13.52/20.95**	+12% / +28% / +4%
	100	34.05/12.75/21.12	35.19/11.44/19.89	**39.75/14.11/21.69**	+13% / +11% / +3%
WikiHow	10	23.95/6.54/15.33	28.34/8.16/19.72	**29.72/9.76/21.02**	+5% / +20% / +7%
	100	25.24/7.52/17.79	31.00/9.68/21.50	**32.73/11.68/24.55**	+6% / +21% / +14%
BIGPATENT	10	28.87/8.30/19.71	-	**37.39/11.44/22.76**	+30% / +38% / +15%
	100	33.52/10.82/22.87	-	**38.66/12.17/23.14**	+15% / +12% / +1%
Xsum	10	19.39/3.45/14.02	-	**32.35/11.86/25.33**	+67% / +244% / +81%
	100	**39.07/16.44/31.27**	-	35.54/13.94/27.79	-9% / -15% / -11%
CNN/DailyMail	10	37.25/15.84/**33.49**	-	**39.34/16.53**/25.40	+6% / +4% / -24%
	100	**40.28/18.21/37.03**	-	39.94/16.96/26.09	-1% / -7% / -30%

Natural Language Processing

Summary and Future Work

❖ Summary

- 문서 요약과 XAI에서의 문서 요약의 역할
- Open Domain QA에서의 정답 추론 근거 활용
- 질의어 중심의 생성 요약
- Meta-Learning 기반의 생성 요약

❖ Future Work

- Multi-Document Summarization(MDS)으로의 확대 연구
- 거대 언어 모델(LLM)을 활용한 질의응답 시스템 적용 및 개선 방안

Thank you for your attention!

고 영 중

(http://nlp.skku.edu, yjko@skku.edu)

SKKU NLP LAB

SKK UNIVERSITY