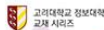


자연어 질의응답 시스템



Natural language processing Bible

**자연어처리
바이블** 핵심이론
응용시스템
딥러닝

임희석 | 고려대학교 자연어처리연구실 지

Human Science
휴먼사이언스

고려대학교 컴퓨터학과
NLP & AI Lab.

Human-inspired AI
임희석



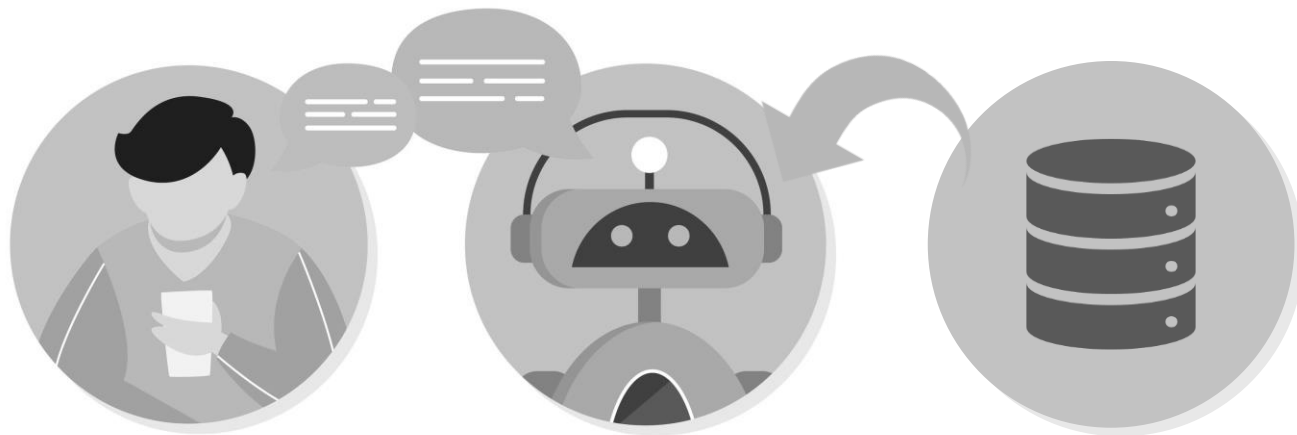
KOREA UNIVERSITY

질의응답 시스템

Question Answering System

1. 개념

- 자연어 질의에 대한 자연어 답변을 보유한 데이터 내에서 찾아 제시해주는 시스템



질의응답 시스템

Question Answering System

1. 개념

- 정보검색(IR) 시스템: “책장에서 적절한 책을 찾아주는 역할”
- 질의응답(QA) 시스템: “책을 찾아 페이지까지 펴서 찾는 정보를 제시해주는 역할”

	IR	QA
입력	검색어	자연어 질의
출력	문서 전체	자연어 답변
정보 획득	출력된 문서 탐색	답변 읽기

질의응답 시스템

Question Answering System

1. 개념

- 1) 딥러닝 이전엔 주로 정보검색 기반 질의응답 시스템 이용
- 2) 자연어처리 분야의 기술 발달과 함께 딥러닝 기반 질의응답 시스템이 여러 형태로 응용

정보검색 기반 질의응답 시스템

Information Retrieval based Question Answering System

전통적인 정보검색기반 질의응답 시스템

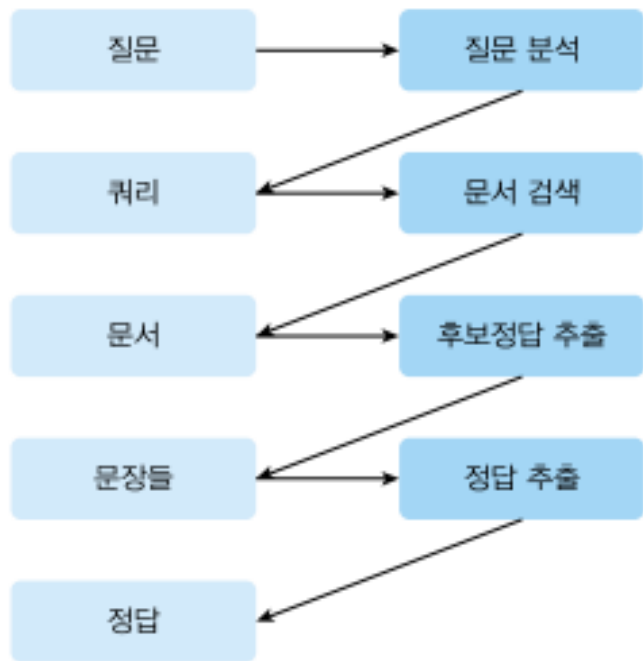


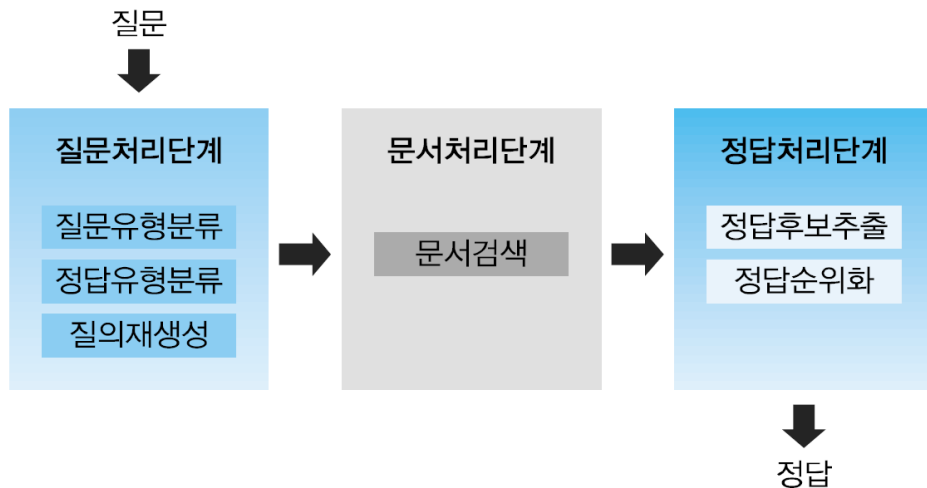
그림 24-1

전형적인 질의응답 처리과정

정보검색 기반 질의응답 시스템

Information Retrieval based Question Answering System

■ 정보검색 기반 질의응답 처리과정



정보검색 기반 질의응답 시스템

Information Retrieval based Question Answering System

1. 질문처리단계(question processing)

- 질문 분석을 통해 질문 유형 분류와 정답 유형 분류를 수행하는 단계

	의문사	예시
1	왜	스티븐 잡스는 왜 죽었나요?
2	누구	거북선을 만든 사람은 누구인가요?
3	어디	고려대학교는 어디 있나요?
4	무엇	에디슨이 첫 번째로 발명한 것은 무엇인가요?
5	무슨	자연어처리는 무슨 용도로 쓰이나요?
6	몇	현재 한국의 피파랭킹은 몇위 인가요?
7	언제	인공신경망은 언제부터 유명해졌나요?
8	얼마나	중국은 얼마나 큰가요?
9	어느	여름이 지나면 어느 계절이 오나요?
10	어떻게	순환신경망은 어떻게 작동하나요?

질문 유형 분류의 예시

의문사	패턴	정답 유형	예문
누구		인물/정의	이순신은 누구인가요?
몇	몇 +(의존) 명사	날짜/시간/기타	틀림은 몇 월에 피나요?
무슨	무슨 + 명사	명사의미에 의존	딸기는 무슨 맛인가요?
어느	어느 + 명사	명사의미에 의존	불국사는 어느 시에 있나요?
무엇	명사 + 무엇 무엇 + 명사	명사의미에 의존	답러님이 무엇인가요?
어디		위치/정의/기타	음성인식을 개발한 회사는 어디인가요?
언제		날짜/기타	독일은 언제 통일 되었나요?
얼마		숫자/기타	백두산은 얼마나 높은가요?
왜		이유	자연어처리는 왜 어렵나요?
어떠하 어찌하		방법/정의/이유	답러님은 어떻게 작동하나요?

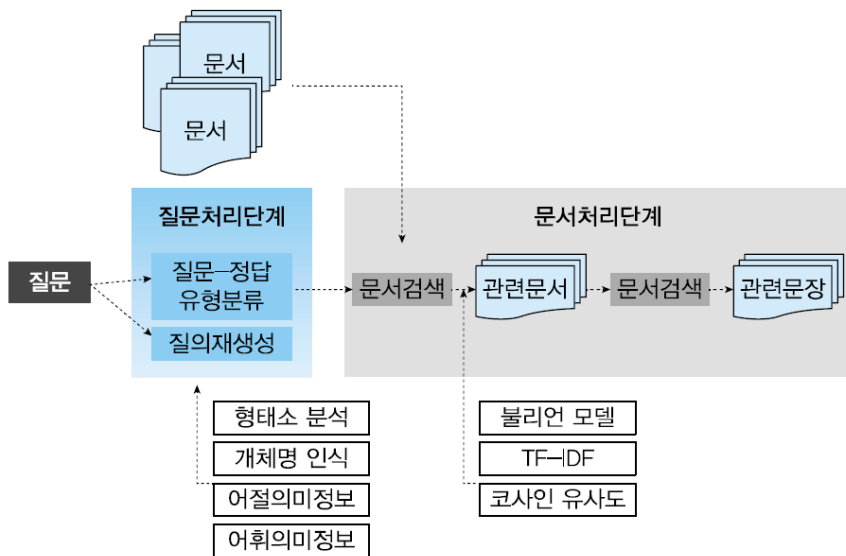
의문사 별 정답 유형 분류의 예시

정보검색 기반 질의응답 시스템

Information Retrieval based Question Answering System

2. 문서처리단계(document processing)

- 정답을 포함하거나 관련성이 높은 문서 혹은 문장을 검색하는 단계



문서처리단계 - 계속

- 문서검색 : 불리언(Boolean) 모델
- 불리언 모델은 질의의 각 단어가 수집된 각 문서에서 출현 했는가의 여부를 표시
- 일치성 외에 어떤 문서가 더 중요한지, 질의어와 더욱 일치하는지 측정 불가

표 11-4 Term-document matrix

	지적인	사업	고속도로	나는	집	학교	컴퓨터	딥러닝	공부하다
문서1	1	0	1	0	0	1	0	0	0
문서2	1	1	1	0	0	0	1	0	1
문서3	0	0	1	0	1	0	0	0	1

문서처리단계 - 계속

- 문서검색 : 벡터 공간 모델
- 불리언 모델의 단점을 해소
- 각 문서에 나타나는 단어들에 가중치를 부여하고, 그 가중치에 따라 질의어와 문서의 유사도를 측정함
- 질의어와 문서의 유사도 혹은 문서간의 유사도를 측정하는 방법에는 자카드 유사도, 코사인 유사도, TF-IDF 등이 주로 사용됨

문서처리단계 - 계속

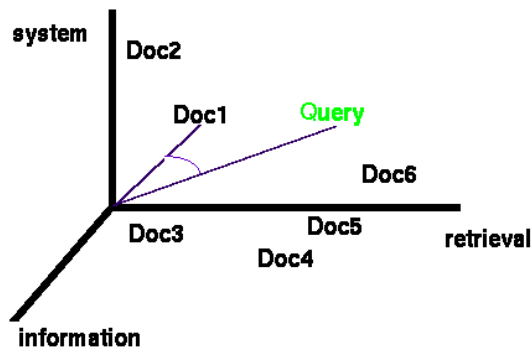
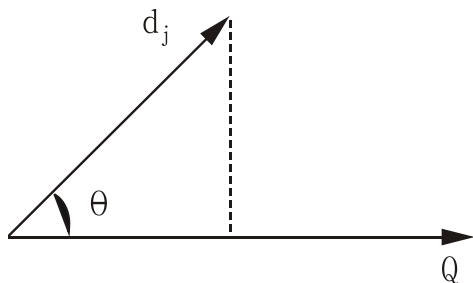
▪ 유사도 측정 방법

- 자카드 유사도는 **A**와 **B** 두개의 집합이 있다고 했을 때, 두 집합의 합집합에서 교집합의 비율을 구한다면 두 집합 **A**와 **B**의 유사도를 구할 수 있다는 아이디어에서 출발함

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (11.1)$$

$$J(Question_1, Docu_1) = \frac{Question_1 \cap Docu_1}{Question_1 \cup Docu_1} \quad (11.2)$$

Cosine Similarity : $\cos(A, B)$



COSINE SIMILARITY

- Definition

$$\vec{q} = (w_{1q}, w_{2q}, \dots, w_{tq}) \quad w_{iq} \geq 0$$

$$\vec{d}_j = (w_{1j}, w_{2j}, \dots, w_{tj}) \quad w_{ij} \geq 0$$

$$\vec{q} \bullet \vec{d} = |\vec{q}| \times |\vec{d}| \times \cos \theta$$

$$\text{sim}(d_i, q) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} = \frac{\sum_{i=1}^t w_{ij} \times w_{iq}}{\sqrt{\sum_{i=1}^t w_{ij}^2} \times \sqrt{\sum_{i=1}^t w_{iq}^2}}$$

$$0 \leq \text{sim}(d_i, q) \leq 1 \quad (\text{cosine similarity})$$

$|\vec{q}|$: Does not affect the ranking

$|\vec{d}_j|$: Normalization in the space of the documents

COSINE SIMILARITY를 이용한 유사도 계산

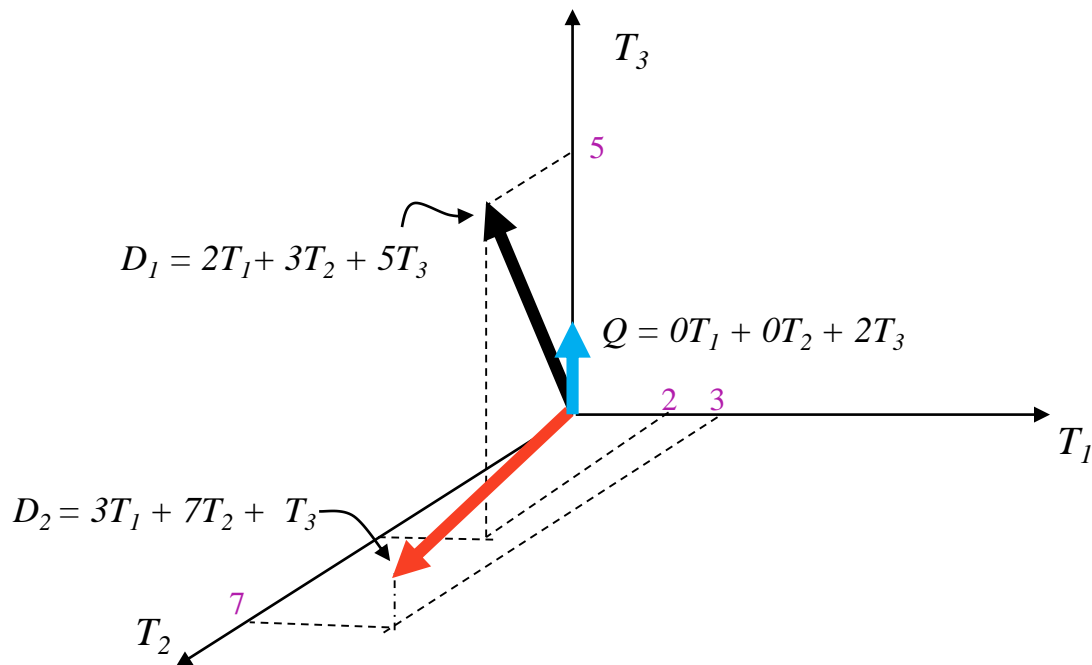
예 2)

$$D_1 = 2T_1 + 3T_2 + 5T_3$$

$$D_2 = 3T_1 + 7T_2 + T_3$$

$$Q = 0T_1 + 0T_2 + 2T_3$$

- D_1 혹은 D_2 중 어느 것이 Q 에 더 유사한가?



문서처리단계 - 계속

▪ 유사도 측정 방법

- 자카드 유사도는질의어와 문서 사이의 유사도를 측정하기에 매우 간단한 방법임
- 하지만, 단어 빈도수(TF)와 문서 빈도수(DF)를 고려하지 않는 단점이 있음
- TF는 “단어가 그 문서에 나타난 횟수”를 나타냄
- 따라서, 문서 내에서 해당 단어의 등장 횟수를 통하여 각 단어에 대한 중요도를 알 수 있음

표 11-5 TF로 나타낸 Term-document matrix

	지적인	사업	고속도로	나는	집	학교	컴퓨터	딥러닝	공부하다
문서1	120	0	20	0	0	39	0	0	0
문서2	30	1	8	0	0	0	31	0	401
문서3	0	0	3	0	37	0	0	0	22

문서처리단계 - 계속

▪ 유사도 측정 방법

- TF와 비슷한 개념으로써 DF는 “해당 단어가 나타난 문서의 수”
- DF가 높은 단어는 많은 문서에서 나타나는 것이므로, 검색에서 별로 중요한 단어가 아님
- IDF는 전체 단어수를 해당 단어의 DF로 나눈 뒤 로그를 취한 값임
- TF-IDF는 TF와 IDF의 곱으로 이루어지며, 특정 문서에서만 자주 등장하는 단어는 중요도가 높다고 판단함

	Doc1	Doc2	Doc3		Doc1	TF	DF	IDF	TF-IDF
Word1	5	0	0		Word1	5	1	$\log(3/2)$	$5\log(3/2)$
Word2	1	0	0		Word2	1	1	$\log(3/2)$	$\log(3/2)$
Word3	5	5	5		Word3	5	3	$\log(3/4)$	$5\log(3/4)$
Word4	3	3	3		Word4	3	3	$\log(3/4)$	$3\log(3/4)$
Word5	3	0	1		Word5	3	2	$\log(3/3)$	0

그림 11-3

TF, IDF, TF-IDF의 관계

정보검색 기반 질의응답 시스템

- Information Retrieval based Question Answering System

3. 정답처리단계(answer processing)

- 검색된 문서 혹은 문장에서 정답 후보에 해당하는 개체, 어휘, 구 등을 추출하는 단계

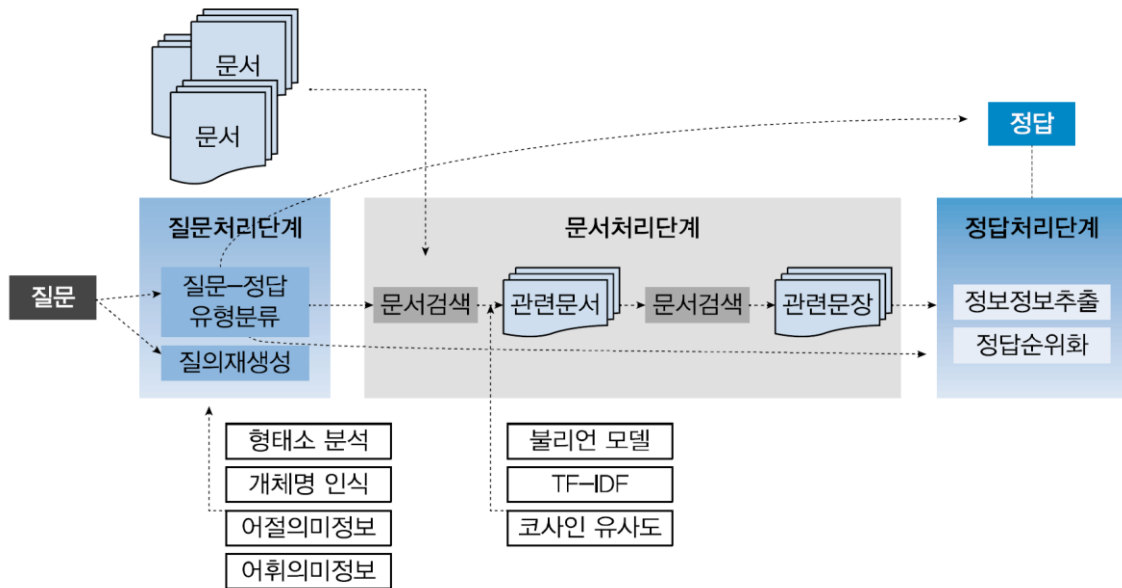


그림 11-4

정답처리단계의 과정

정답처리단계 - 계속

- 주어진 질문에 따른 정답유형에 집중하고, 이에 따라 검색된 관련문서에 대한 형태소 분석, 개체명 태그 분석을 진행
- 그 다음, 정답유형과 같은 개체를 갖는 정답후보를 추출함

질문	인도의 총리는 누구인가?
정답유형	인물, 정의
관련문서	간디 탄생 150주년을 맞아 인도 정부가 독립운동가 마하트마 간디의 흉상을 연세대학교에 선물한다. 간디 흉상 제막식에는 나렌드라 모디 인도 총리가 직접 참석할 예정이다.
정답후보	간디, 마하트마 간디, 나렌드라 모디

그림 11-5 정답후보추출의 예

딥러닝 기반 질의응답 시스템

Deep Learning based Question Answering System

딥러닝 기반 질의응답 시스템

Deep Learning based Question Answering System

1. Machine Reading Comprehension (MRC)
2. Visual Question Answering (VQA)
3. Visual Dialog
4. Commonsense QA
5. Neural-symbolic QA

Machine Reading Comprehension

1. 개념

- 주어진 문서를 빠르게 이해하고 질문에 대한 답을 해당 문서에서 찾아주는 작업

3일 코스피가 전날보다 59.81포인트(2.87%) 오른 2147.00으로 마감했다. 외국인과 기관의 쌍끌이 매수세에 코스피는 지난 2월 25일 이후 처음으로 2100을 넘겨 장을 마감했다. 이날 유가증권시장에서 개인은 1조3276을 순매도했고 외국인은 2094억원, 기관은 1조1583억원을 순매수했다.

이날 유가증권시장의 거래량은 11억2207만주였고, 거래대금은 16조7754억원으로 역대 최대치를 갈아치웠다. 지금까지는 지난달 28일 14조4792억원이 최대치였다.

코스피 지수는 얼마인가요?



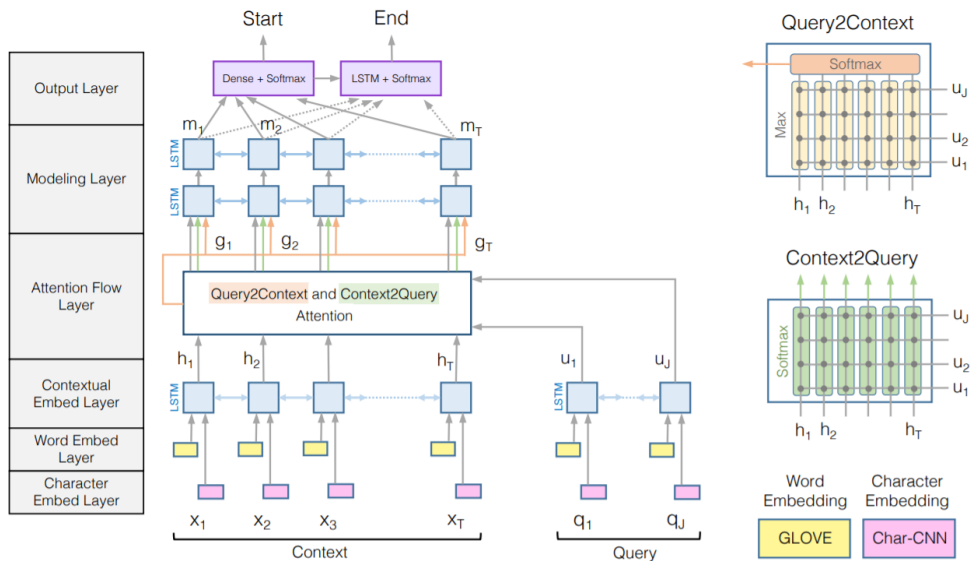
AI 독해 분석 결과

3일 코스피가 전날보다 59.81포인트(2.87%) 오른 **2147.00**으로 마감했다. 외국인과 기관의 쌍끌이 매수세에 코스피는 지난 2월 25일 이후 처음으로 2100을 넘겨 장을 마감했다. 이날 유가증권시장에서 개인은 1조3276을 순매도했고 외국인은 2094억원, 기관은 1조1583억원을 순매수했다. 이날 유가증권시장의 거래량은 11억2207만주였고, 거래대금은 16조7754억원으로 역대 최대치를 갈아치웠다. 지금까지는 지난달 28일 14조4792억원이 최대치였다.

Machine Reading Comprehension

2. 모델 소개

- BiDirectional Attention Flow Model (BiDAF, ICLR 2017)

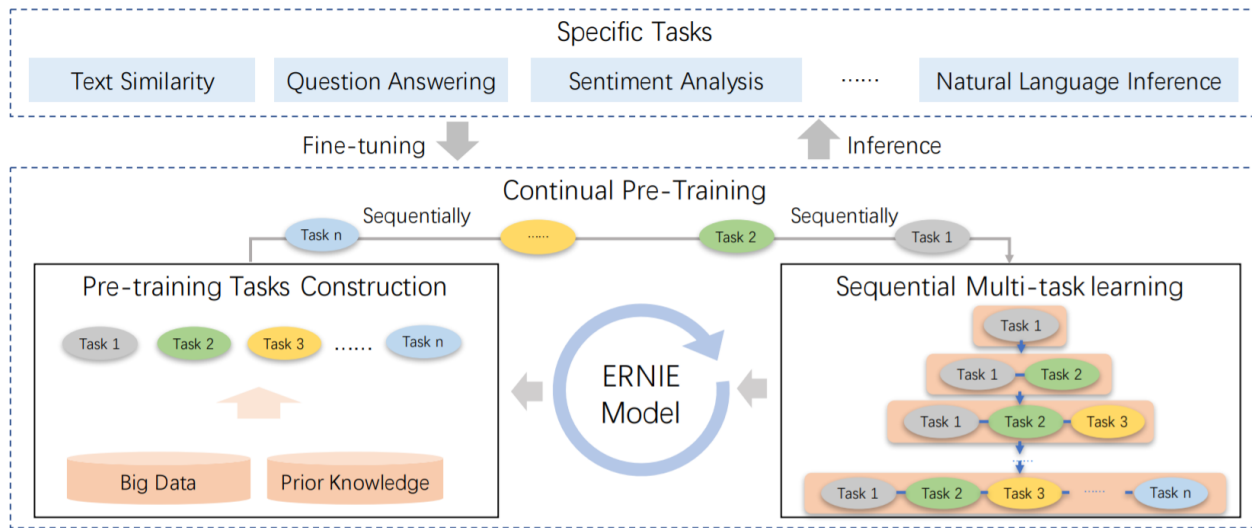


- Seo, Minjoon, et al. "Bidirectional attention flow for machine comprehension." *arXiv preprint arXiv:1611.01603* (2016).

Machine Reading Comprehension

2. 모델 소개

- Enhanced Representation through Knowledge Integration (ERNIE 2.0)



- Sun, Yu, et al. "Ernie 2.0: A continual pre-training framework for language understanding." arXiv preprint arXiv:1907.12412 (2019).

Machine Reading Comprehension

3. KU MRC 시연 영상



Visual Question Answering

1. 개념

- Text-based QA와 다르게 semantic and visual understanding이 함께 요구되는 작업
- 이미지에 근거하는 질문과 정답으로 학습 데이터가 구성됨



What kind of store is this?	bakery	art supplies
	bakery	grocery
	pastry	grocery
Is the display case as full as it could be?	no	no
	no	yes
	no	yes



How many bikes are there?	2	3
	2	4
	2	12
What number is the bus?	48	4
	48	46
	48	number 6

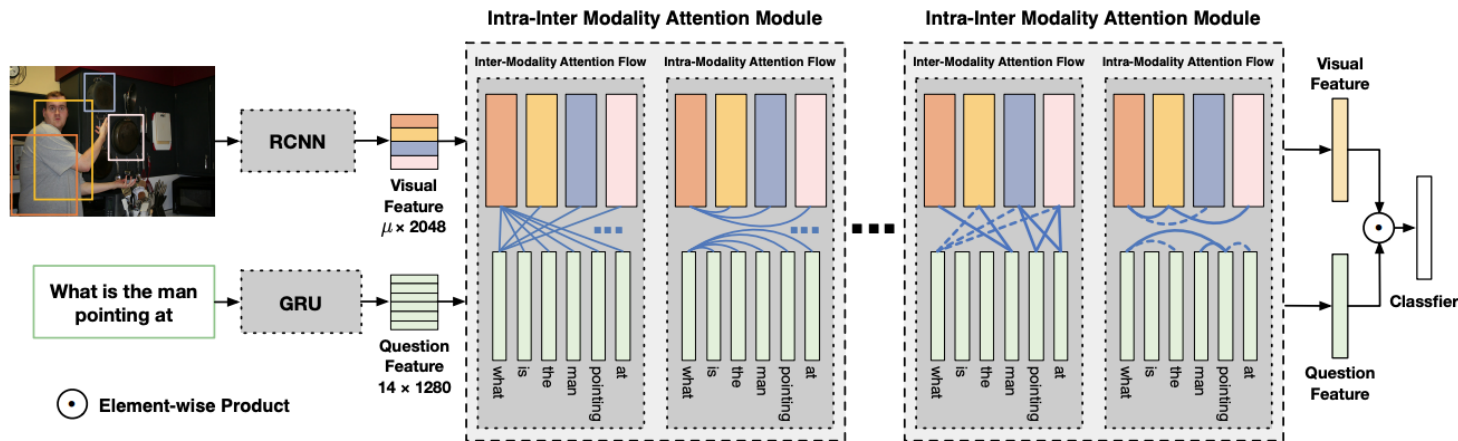


- Antol, Stanislaw, et al. "Vqa: Visual question answering." Proceedings of the IEEE international conference on computer vision. 2015.

Visual Question Answering

2. 모델

- 2019년 이전: 각 modality에서 별도로 특징을 추출하는 attention based approach 유행
- Bottom-up, MFH, DCN, Counter, BAN, DFAF 등



- Peng, Gao, et al. "Dynamic fusion with intra-and inter-modality attention flow for visual question answering." arXiv preprint arXiv:1812.05252 (2018).

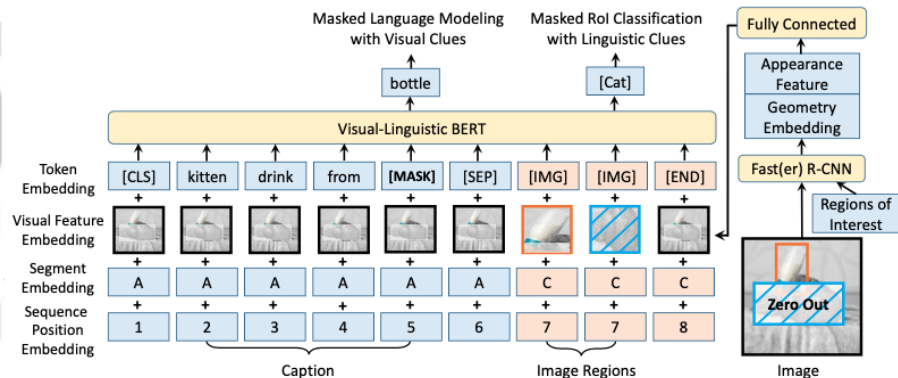
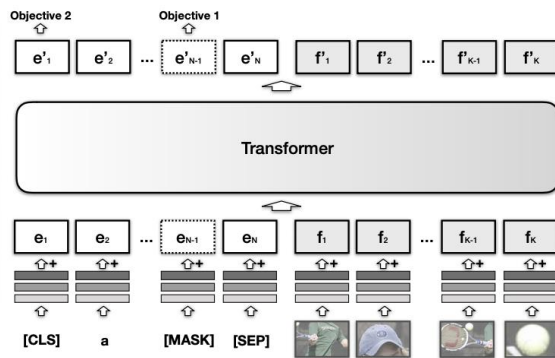
Visual Question Answering

2. 모델

- 2019년 이후: multi-modality를 반영한 pretrained approach 유행
- VisualBERT, VL-BERT, LXMERT 등



A person hits a ball with a tennis racket



- Li, Liunian Harold, et al. "Visualbert: A simple and performant baseline for vision and language." arXiv preprint arXiv:1908.03557 (2019).

Visual Question Answering

2. 모델

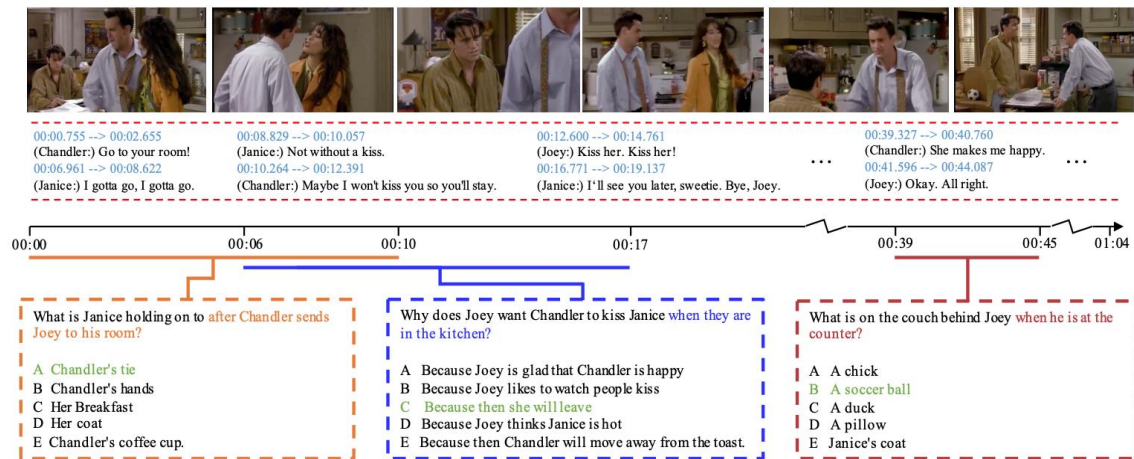
- VQA 2.0에 대한 성능 비교
- Pretrained based approach의 우세

Model	test-dev	test-std
Bottom-up	65.32	65.67
DCN	66.87	66.97
BAN	69.66	N/A
DFAF	70.22	70.34
ViLBERT	70.55	70.92
VisualBERT	70.80	71.00
VL-BERT	71.79	72.22
LXMERT	72.42	72.54

Visual Question Answering

3. VQA의 한계

- Single-turn QA로 이루어져 있어 실제로 human-interactive하다고 할 수 없음
- 최근 multi-modal 확장 연구로 visual dialog, Video QA와 같은 task가 제안됨

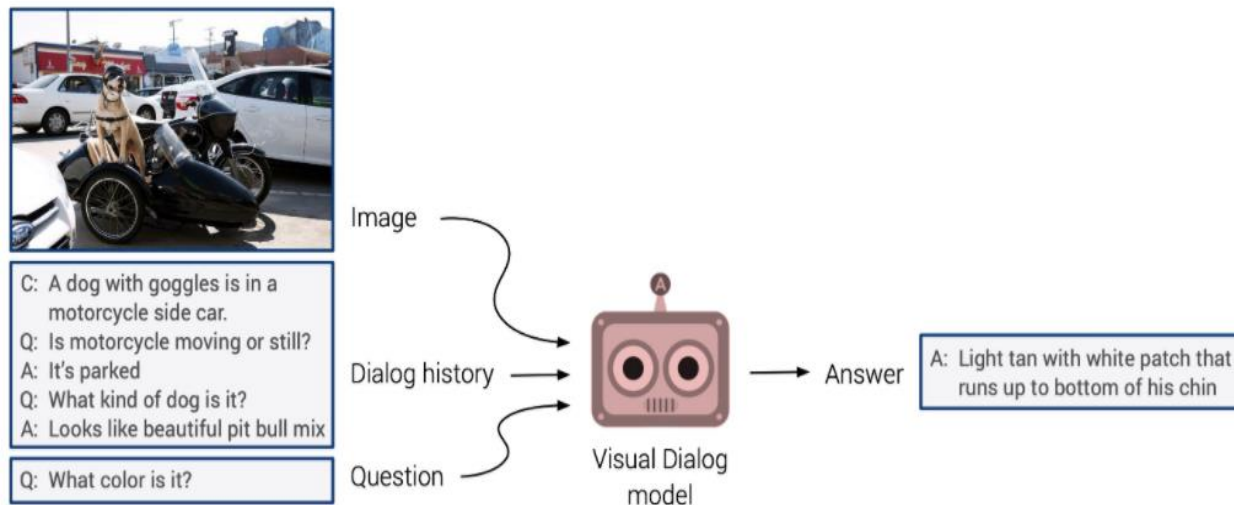


- Lei, Jie, et al. "Tvqa: Localized, compositional video question answering." arXiv preprint arXiv:1809.01696 (2018).

Visual Dialog

1. 개념

- VQA와는 다르게 하나의 이미지와 multi-turn QA dialog로 구성된 task



- Das, Abhishek, et al. "Visual dialog." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

Visual Dialog

1. 개념

- Dialog turn에 따라 topic이 변하는 문제
- Visual reference resolution 문제

Dialog Topics

People Food household goods



Cap: 2 small **kids** eating large
carrots on a **bed**

Q1: is this in color?

A1: yes

Q2: is it a big or little **bed**?

A2: there is no **bed** **they** are sitting
on a **blanket** on the floor

Q3: what color is it the **blanket**?

A3: multicolored blues

Q4: are the **kids** **boys** or girls?

A4: **boys**

Q5: how old do **they** look?

A5: 7-9

Q6: do **they** have any other **snacks**?

A6: no

Q7: is there a **tv** in the room?

A7: not that i see

Q8: are **they** smiling?

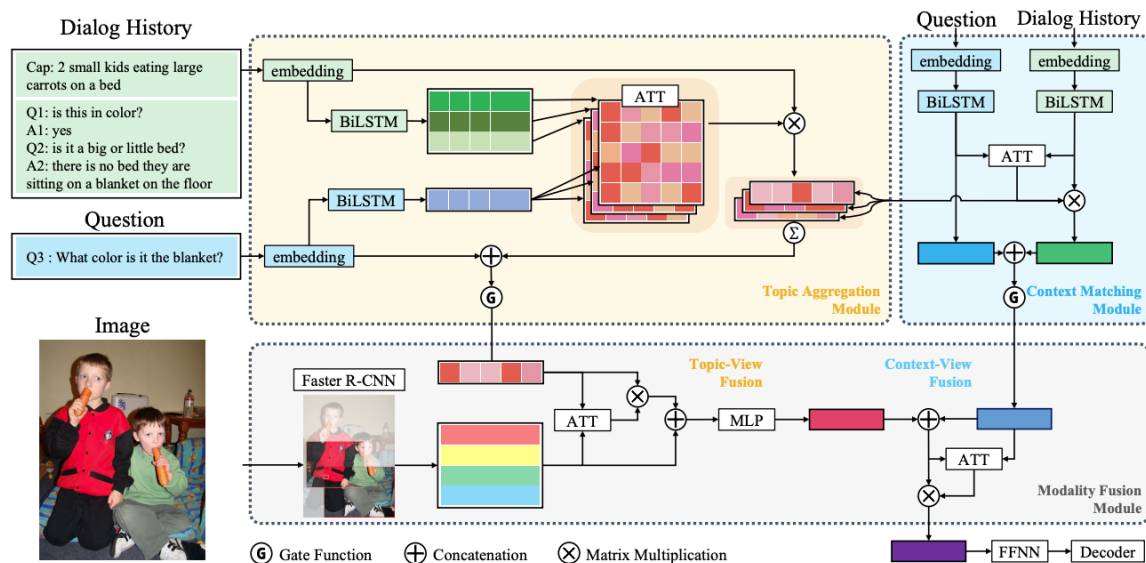
A8: no

- Park, Sungjin, et al. "Multi-View Attention Networks for Visual Dialog." arXiv preprint arXiv:2004.14025 (2020).

Visual Dialog

2. 모델 소개

- Multi-View Attention Network (MVAN)



- Park, Sungjin, et al. "Multi-View Attention Networks for Visual Dialog." arXiv preprint arXiv:2004.14025 (2020).

Visual Dialog

2. 모델 소개

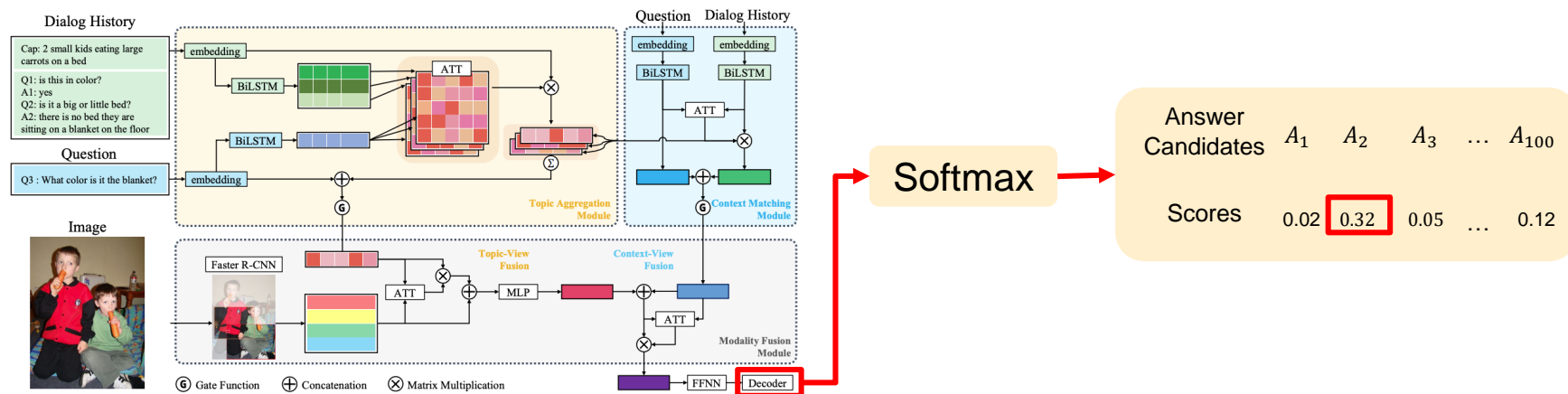
- Attention-based / Graph-based / Pretraining-based Approaches

Model	VisDial v1.0 (test-std)						VisDial v0.9 (val)				
	NDCG	MRR	R@1	R@5	R@10	Mean	MRR	R@1	R@5	R@10	Mean
LF	45.31	55.42	40.95	72.45	82.83	5.95	58.07	43.82	74.68	84.07	5.78
HRE	45.46	54.16	39.93	70.45	81.50	6.41	58.46	44.67	74.50	84.22	5.72
MN	47.50	55.49	40.98	72.30	83.30	5.92	59.65	45.55	76.22	85.37	5.46
HCIAE	-	-	-	-	-	-	62.22	48.48	78.75	87.59	4.81
AMEM	-	-	-	-	-	-	62.27	48.53	78.66	87.43	4.86
CoAtt	-	-	-	-	-	-	63.98	50.29	80.71	88.81	4.47
CorefNMN	54.70	61.50	47.55	78.10	88.80	4.40	64.10	50.92	80.18	88.81	4.45
RVA	55.59	63.03	49.03	80.40	89.83	4.18	66.34	52.71	82.97	90.73	3.93
DualVD	56.32	63.23	49.25	80.23	89.70	<u>4.11</u>	62.94	48.64	80.89	89.94	4.17
HACAN	57.17	<u>64.22</u>	<u>50.88</u>	80.63	89.45	4.20	67.92	54.76	83.03	90.68	3.97
Synergistic	57.32	62.20	47.90	80.43	89.95	4.17	-	-	-	-	-
DAN	<u>57.59</u>	63.20	49.63	79.75	89.35	4.30	66.38	53.33	82.42	90.38	4.04
MVAN	59.37	64.84	51.45	81.12	90.65	3.97	67.65	54.65	83.85	91.47	3.73
FGA	52.10	63.70	49.58	<u>80.98</u>	88.55	4.51	67.12	54.02	83.21	90.47	4.08
CAG	56.64	63.49	49.85	80.63	<u>90.15</u>	<u>4.11</u>	67.56	54.64	<u>83.72</u>	91.48	<u>3.75</u>
SGLN	60.77	58.40	44.15	75.65	85.70	5.22	-	-	-	-	-
VisDial-BERT	74.47	50.74	37.95	64.13	80.00	6.28	-	-	-	-	-

Visual Dialog

3. Visual Dialog 한계점

- Discriminative decoder



- Park, Sungjin, et al. "Multi-View Attention Networks for Visual Dialog." arXiv preprint arXiv:2004.14025 (2020).

Visual Dialog

3. Visual Dialog 한계점

- 주관적이고 모호한 질문/대답



Q2: how old is he?

GT: looks like maybe 26 (20)

Prediction: 1) 20s 2) he looks like he is a teen
3) 20's



Q7: is the weather sunny?

GT: it appears overcast (23)

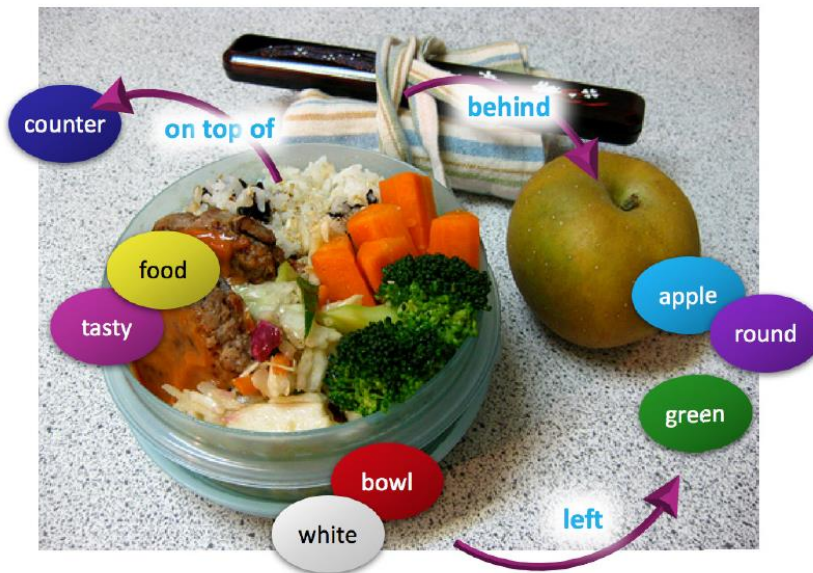
Prediction: 1) yes but it is blue sky 2) yes very clear
3) yes it's sunny

- Park, Sungjin, et al. "Multi-View Attention Networks for Visual Dialog." arXiv preprint arXiv:2004.14025 (2020).

Visual Dialog

3. Future Work

- Generative decoder 연구
- 외부 Knowledge 사용



- Hudson, Drew A., and Christopher D. Manning. "Gqa: A new dataset for real-world visual reasoning and compositional question answering." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

Commonsense QA

1. 개념

- 관련된 문서 없이도 일반상식(commonsense)을 이용하여 질의응답을 수행하는 task

Where on a **river** can you hold a cup upright to catch water on a sunny day?

✓ **waterfall**, ✗ **bridge**, ✗ **valley**, ✗ **pebble**, ✗ **mountain**

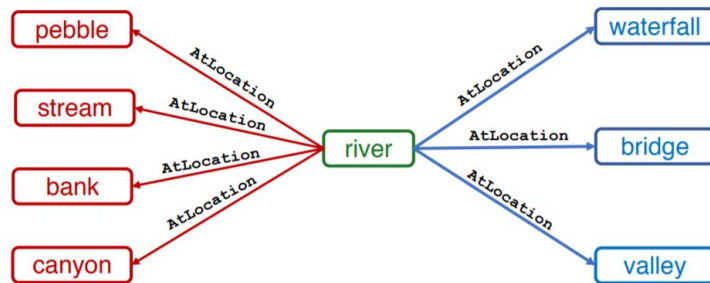
Where can I stand on a **river** to see water falling without getting wet?

✗ **waterfall**, ✓ **bridge**, ✗ **valley**, ✗ **stream**, ✗ **bottom**

I'm crossing the **river**, my feet are wet but my body is dry, where am I?

✗ **waterfall**, ✗ **bridge**, ✓ **valley**, ✗ **bank**, ✗ **island**

CommonsenseQA 데이터셋 예시



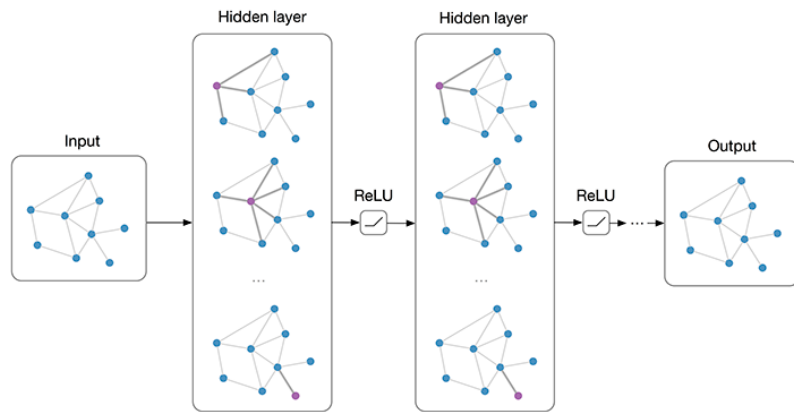
ConceptNet의 subgraph 예시

- Talmor, Alon, et al. "Commonsenseqa: A question answering challenge targeting commonsense knowledge." arXiv preprint arXiv:1811.00937 (2018).

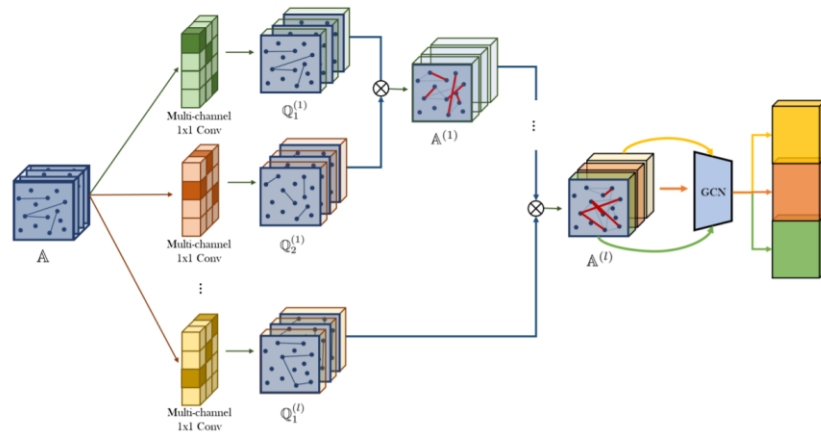
Commonsense QA

2. 모델 소개

- 지식 그래프를 인코딩하기 위한 네트워크



Graph Convolutional Networks (GCNs)



Graph Transformer Networks (GTNs)

- Yun, Seongjun et al., "Graph Transformer Networks", Neural Information Processing Systems (NeurIPS), 2019

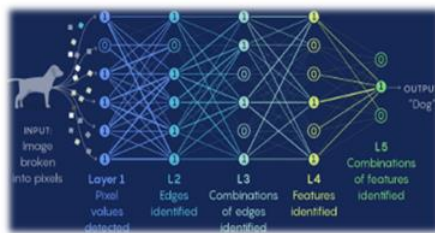
Commonsense QA

3. 한계점

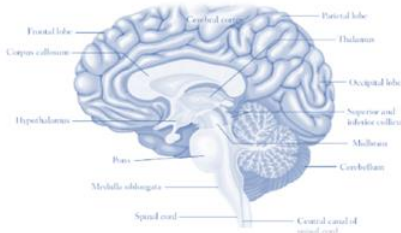
- 구축된 지식의 반영은 가능하나, 학습 과정에서 새로운 지식을 생성/저장할 수 없음
- 추론 과정에 대한 설명력이 부족함
- 딥러닝 모델 내에 지식을 학습시킬 경우 지식과 연산 과정이 혼재하는 문제 발생



암묵적 지식 유추



추론 처리와 설명력 부족

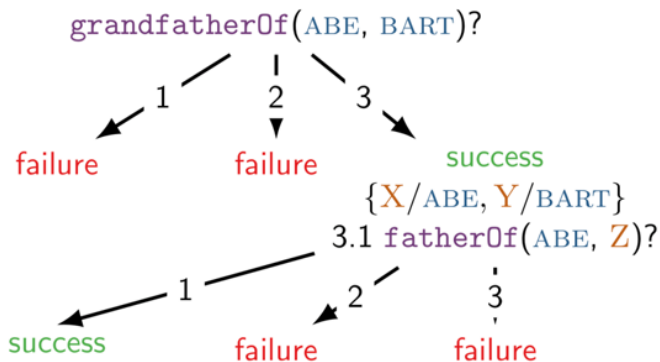
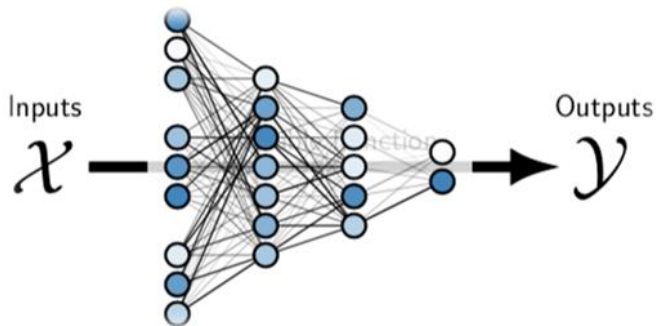


기억 공간의 한계

Neural-symbolic QA

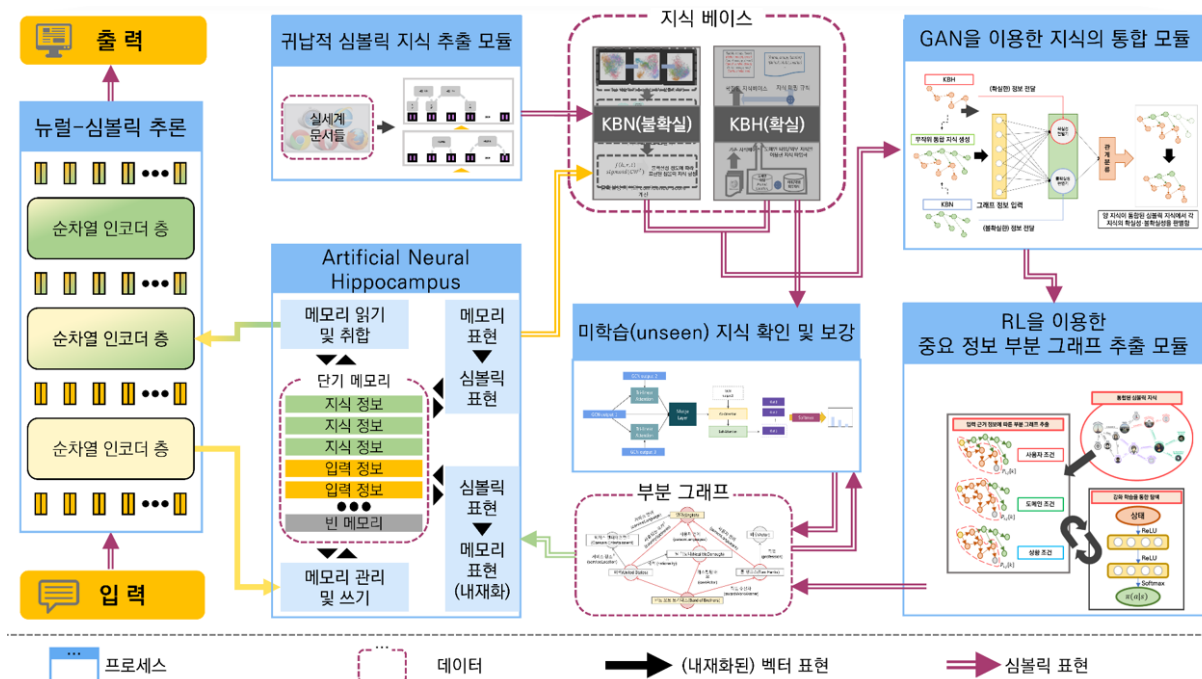
1. 개념

- 질의응답을 위해 지식을 반영할 뿐만 아니라, 지식을 생성하고, 추론 과정까지 설명 가능한, 딥러닝 네트워크와 심볼 지식의 통합을 요구하는 작업



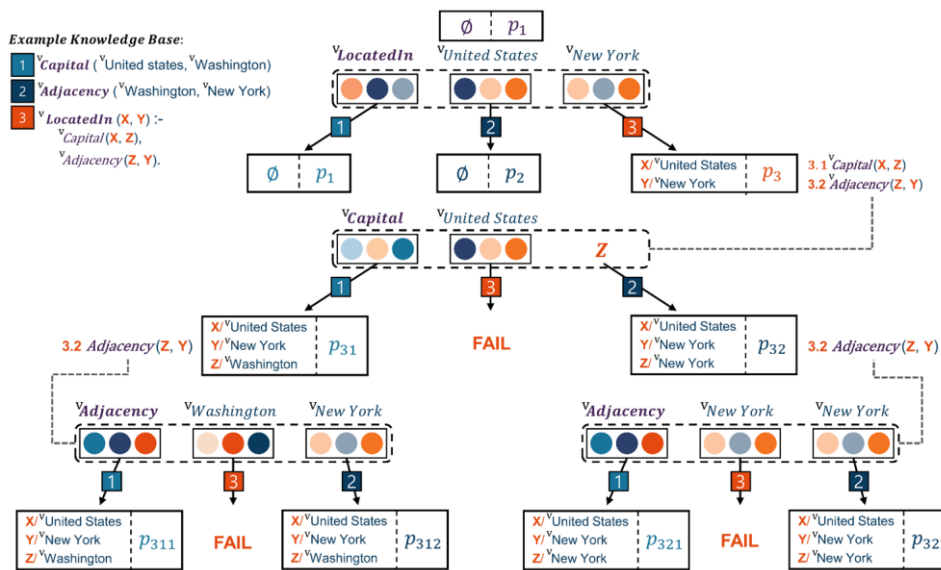
Neural-symbolic QA

2. 모델 소개: Neural-symbolic AI (2020)



Neural-symbolic QA

2. 모델 소개: Differentiable Reasoning (2019)



Neural-symbolic QA

3. Future Work

- 딥러닝 네트워크의 학습 능력과 심볼릭 지식을 통합한 질의응답 시스템