



Machine Reading Comprehension Survey

Daumsoft AI Lab
KwangJune Choi



MRC (Machine Reading Comprehension)

기계독해란?

주어진 문서와 문서에 대한 질문이 주어졌을 때, 문서 내에서 최적의 답안을 찾아내는 기술을 말함

MRC (Machine Reading Comprehension)

MRC의 지향점은 **Question Answering** system (or service)

기존 QA 서비스는 DB에서 데이터를 찾고 키워드를 기반으로 한 리스트를 제공하는 방식

MRC를 활용한 QA는 이러한 방식을 넘어 정형 데이터든 비정형 데이터든 다양한 데이터 속에서 키워드 기반이 아닌 사용자의 의도를 분석해 원하는 답을 찾아주는 것을 목표로 함

MRC (Machine Reading Comprehension)

QA는 크게 두 가지로 구분

1. Open-domain QA

거의 모든 것에 대한 질문을 다루며, general ontologies와 범용적 지식에 의존함.
위키피디아가 활용이 예시. Facebook research의 DrQA가 대표적인 ODQA System.

2. Closed-domain QA

특정 도메인(의약품 or 법률 등)에 해당하는 질문을 다룸. 데이터베이스에 적합한 모델을 사용하여 도메인 별 지식을 이용가능. cdQA-suite가 대표적.

MRC (Machine Reading Comprehension)

IR-based QA & KB QA

- Information Retrieval Question Answering
 - 검색 기반 질의 응답.
 - 정보 검색을 통해 해당 질의에 대해 정답을 포함하는 문서를 찾는 정보 검색 과정을 거친 후 검색된 문장 내에서 질의에 대한 정답을 추출.
- Knowledge base Question Answering
 - 질의에서 묻고자 하는 토픽 개체를 인식하고 토픽 개체와 질의 문장과의 관계 추출 과정을 거친 후 질문의 해당 정답을 지식 기반(Knowledge Base)상의 트리플 내에서 찾는 방식.

MRC (Machine Reading Comprehension)

- 초기 MRC 시스템은 1970년대 Lehner의 QUALM system에서 처음 제안

(작은 데이터 사이즈와 특정 도메인 한계로 인해 1980년대와 1990년대 널리 적용되지 못함)

- 1999년 Hirschman과 연구진은 5개의 “wh”(what, where, when, why and who)질문과 함께 3부터 6등급의 내용이 담긴 데이터세트 발표, 이후 MRC 연구가 부상

(하지만 당시 MRC task를 해결하는 연구는 주로 rule-based와 machine-learning 기반으로 이루어짐)

MRC (Machine Reading Comprehension)

당시 MRC task를 해결하는 연구는 주로 rule-based와 machine-learning 기반

질문과 context가 있는 문장을 단어 집합으로 나타내는 bag-of-words 기법이 주를 이루었고, context와 context 사이에서 나타나는 모든 단어들이 정답으로 설정

- 이후 Riloff와 연구진은 Quarc라는 rule-based 기반의 MRC 시스템을 구축
- Poon과 연구진은 machine-learning method에 bootstrapping, Markov logic 그리고 self-supervised learning을 적용

MRC (Machine Reading Comprehension)

앞서 제시된 방법들의 한계성

1. Mainly based on hand-crafted rules나 features를 사람이 직접 구현하기 때문에 cost 소모가 큼
2. 다양한 데이터에 일반화가 힘들
3. large-scaled datasets에서 제대로 효과를 보기 힘들
4. long-range dependencies가 무시됨
5. contextual information이 제대로 반영 안됨

MRC (Machine Reading Comprehension)

MRC 연구는 2015년도부터 새로운 국면을 맞이함

1. Deep learning 기반의 MRC 연구 등장

(contextual information을 포착하는데 우수한 성능을 보이게 됨)

2. Large-size Datasets 등장

(CNN&Daily Mail, SQuAD, MS MARCO, RACE etc..)

MRC (Machine Reading Comprehension)

MRC Datasets

<i>datasets</i>	<i>build</i>	<i>questions</i>	<i>docs</i>
CNN/Daily Mail	generate fill in blank question	387,420(CNN)/ 997,467(Daily Mail)	92,579(CNN)/ 219,506(Daily Mail)
SQuAD	crowdsourcing	107,785	563(Wikipedia)
MS MARCO	big search log	100,000	1M answer paragraph/ 200,000 + webdocs links
NewsQA	crowdsourcing	119,633	12,744(CNN)

SQuAD 2.0

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Apr 06, 2020	SA-Net on Albert (ensemble) QIANXIN	90.724	93.011
2 Apr 05, 2020	Retro-Reader (ensemble) Shanghai Jiao Tong University http://arxiv.org/abs/2001.09694	90.578	92.978
3 Mar 12, 2020	ALBERT + DAAF + Verifier (ensemble) PINGAN Omni-Sinitic	90.386	92.777
4 Jan 10, 2020	Retro-Reader on ALBERT (ensemble) Shanghai Jiao Tong University http://arxiv.org/abs/2001.09694	90.115	92.580
5 Nov 06, 2019	ALBERT + DAAF + Verifier (ensemble) PINGAN Omni-Sinitic	90.002	92.425
6 Sep 18, 2019	ALBERT (ensemble model) Google Research & TTIC https://arxiv.org/abs/1909.11942	89.731	92.215
6 Feb 25, 2020	Albert_Verifier_AA_Net (ensemble) QIANXIN	89.743	92.180
7 Mar 28, 2020	Retro-Reader on ELECTRA (single model) Shanghai Jiao Tong University http://arxiv.org/abs/2001.09694	89.562	92.052
7 Mar 27, 2020	albert+KD+transfer (ensemble) Anonymous	89.461	92.134
8 Apr 08, 2020	ALBERT-LSTM (ensemble) oppo.tensorlab	89.224	91.853
8 Apr 21, 2020	albert+KD+transfer+twopass (single) SPPD	89.111	91.877

Rank	Reg. Date	Model	EM	F1
-	2019.09.05	Human Performance	68.82	83.86
1	2020.04.29	SDS-NET (single model) Sanghwan Bae & Soonhwan Kwon	73.51	86.56
2	2020.04.20	LaRva (single model) NAVER Clova AI LaRva	66.95	83.54
3	2020.02.21	CNS-BERT (single model) Seungyoung Lim	68.39	82.62
4	2020.03.10	CNS-BERT_light (single model) Seungyoung Lim	66.96	80.48
5	2020.01.29	onepick-open-simple (single model) SeoHyeong Jeong	58.21	77.33
6	2019.12.23	HanBert-54k-N (single model) TwoBlock Ai	55.28	71.19
7	2019.12.24	HanBert-90k-N (single model) TwoBlock Ai	54.28	70.44
8	2019.11.28	KNU-baseline (single model) Kangwon National University, Intelligent Software Lab.	26.53	66.69

KorQuAD 2.0

MRC
Performance_20200508

MRC (Machine Reading Comprehension)

MRC Processing

1. 기계가 문서와 질문을 읽음
2. 추론 (답이 나올 만한 pattern을 학습)
3. 글에서 정답을 찾음 (위치)

MRC (Machine Reading Comprehension)

ex) BiDirectional Attention Flow(Seo et al., 2017) - Minjoon Seo

1. 문서의 단어와 질문의 단어를 가져옴
2. Glove, char-CNN
3. BiDirectional RNN
4. Co Attention Layer, Q2A, C2Q 두 개의 Attention 사용

MRC (Machine Reading Comprehension)

MRC Tasks

(현재 연구와 서비스 되고 있는 MRC Task는 크게 네 가지로 구분할 수 있고, Task에 부합되는 데이터 세트를 정리하면 다음과 같음)

1. Cloze Tests (CNN&Daily Mail)
2. Multiple Choice (RACE)
3. Span Extraction (SQuAD)
4. Free Answering (MS MARCO)

MRC (Machine Reading Comprehension)

Cloze Tests

- 문맥을 파악하여 빈칸의 word나 entity를 채워넣는 방식

대표 데이터 세트 (**CNN & Daily Mail**)

- Google Deepmind는 2007년 4월부터 2015년 4월까지 미국 CNN 뉴스기사와 2010년 6월부터 2015년 4월까지 영국 Daily Mail 뉴스 기사를 각각 수집함
- 정답이 들어있는 문서 셋은 총 312,085건

MRC (Machine Reading Comprehension)

Multiple Choice

- 다지선다형

대표 데이터세트 (**RACE**)

- 중국 영어 시험에서 수집된 데이터세트
- 28,000 이상의 passages와 약 100,000개의 question이 포함되어 있음

MRC (Machine Reading Comprehension)

Span Extraction

- 문맥에서 정답에 해당하는 문맥 내 한 범위를 추출.

대표 데이터세트 (**SQuAD**)

- Stanford에서 2016년 Wikipedia 문서 536건을 대상으로 107,785 건의 질문을 크라우드 소싱을 통해 만듦.

MRC (Machine Reading Comprehension)

Free Answering

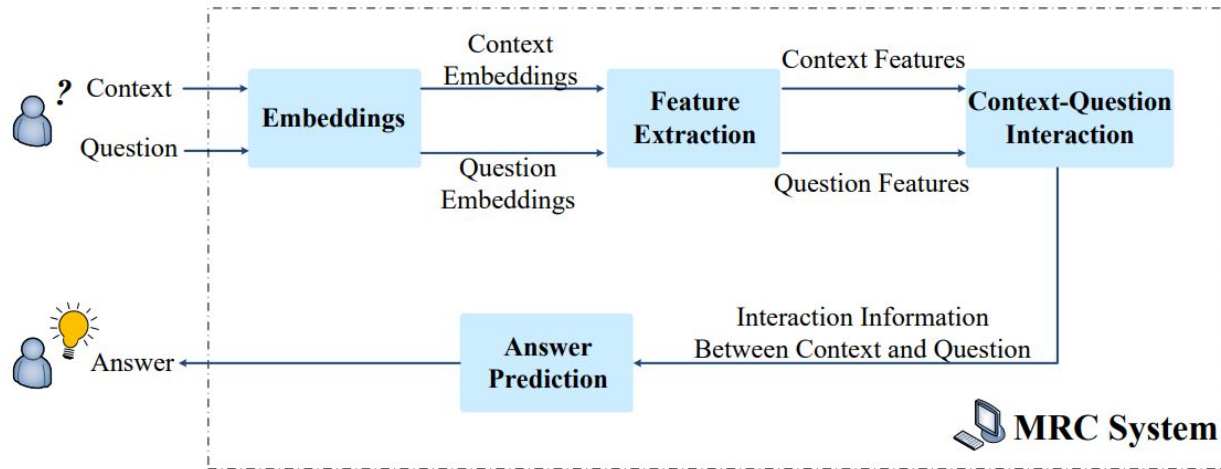
- 문맥의 한 범위라는 제한된 답을 주는 Span extraction의 한계를 극복
- 타 task에 비해 완성도가 높고 답변 형식에 자유도가 높음

대표 데이터세트 (**MS MARCO**)

- 사용자가 실제 질문하는 문장과 정답 단락이 다르게 표현되는 데이터세트 구축
- Bing search engine user log로부터 100,000건의 질문을 구축

Classic Neural MRC Models

Models	Methods			
	Embeddings	Feature Extraction	Context-Question Interaction	Answer Prediction
Attentive Reader (Hermann et al. 2015) [23]	Conventional	RNN	Unidirectional One-Hop	Word Predictor
Impatient Reader (Hermann et al. 2015) [23]	Conventional	RNN	Unidirectional Multi-hop	Word Predictor
End-to-End Memory Networks (Sukhbaatar et al. 2015) [80]	Conventional	/	Multi-Hop	Word Predictor
Stanford Reader (Chen et al. 2016) [6]	Conventional	RNN	Unidirectional One-Hop	Word Predictor
AS Reader (Kadlec et al. 2016) [33]	Conventional	RNN	Unidirectional One-Hop	Word Predictor
IA Reader (Sordani et al. 2016) [78]	Conventional	RNN	Unidirectional Multi-Hop	Word Predictor
Match-LSTM (Wang & Jiang 2016) [91]	Conventional	RNN	Unidirectional Multi-Hop	Span Extractor
Bi-DAF (Seo et al. 2016) [75]	Multiple Granularity	RNN	Bidirectional One-Hop	Span Extractor
DCN (Xiong et al. 2016) [100]	Conventional	RNN	Bidirectional One-Hop	Span Extractor
AoA Reader (Cui et al. 2017) [13]	Conventional	RNN	Bidirectional One-hop	Word Predictor
GA Reader (Dhingra et al. 2017) [18]	Conventional	RNN	Unidirectional Multi-Hop	Word Predictor
MEMEN (Pan et al. 2017) [57]	Multiple Granularity	RNN	Multi-Hop	Span Extractor
R-NET (Wang et al. 2017) [94]	Multiple Granularity	RNN	Unidirectional Multi-Hop	Span Extractor
DCN+ (Xiong et al. 2017) [101]	Conventional	RNN	Bidirectional Multi-Hop	Span Extractor
S-NET (Tan et al. 2017) [84]	Multiple Granularity	RNN	Unidirectional Multi-Hop	Answer Generator
CoVe + DCN (McCann et al. 2017) [47]	Contextual	RNN	Bidirectional One-Hop	Span Extractor
QANet (Yu et al. 2018) [106]	Multiple Granularity	Transformer	Bidirectional One-Hop	Span Extractor
Reinforced Mnemonic Reader (Hu et al. 2018) [28]	Multiple Granularity	RNN	Bidirectional Multi-Hop	Span Extractor
CSA (Chen et al. 2018) [8]	Multiple Granularity	RNN&CNN	Bidirectional One-Hop	Option Selector
CNN Model (Chaturvedi et al. 2018) [4]	Conventional	CNN	Unidirectional One-Hop	Option Selector
Hierarchical Attention Flow (Zhu et al. 2018) [111]	Conventional	RNN	Bidirectional One-Hop	Option Selector
ELMo + improved Bi-DAF (Peters et al. 2018) [62]	Contextual	RNN	Bidirectional One-Hop	Span Extractor
GPT (Radford et al. 2018)[65]	Contextual	/	/	Option Selector
BERT (Devlin et al. 2018) [16]	Contextual	/	/	Span Extractor



MRC General Architectures

1. Embedding
2. Feature Extraction
3. Context-Question Interaction
4. Answer Prediction

1. Embedding

기계가 자연어를 직접 이해할 수 없기 때문에 MRC 시스템 시작 시 입력 단어를 고정 길이 벡터로 변경하는 임베딩 필요. context와 question을 input으로 사용하여 context와 question에 대한 embedding 값을 출력

2. Feature Extraction

context와 question을 더 잘 이해하기 위해 contextual information을 추출하는 것을 목표

3. Context-Question Interaction

interaction 추출을 위해 query와 관련된 context의 일부를 강조하는 단방향 혹은 양방향 attention mechanism이 활용

4. Answer Prediction

MRC system의 마지막 요소로 이전 모듈들의 정보 연산을 통해 획득한 final answer를 output으로 출력함. task에 따라서 별개로 적용

MRC (Machine Reading Comprehension)

Attention Mechanisms for MRC Models

- Mutual-matching Attention
- Self-matching Attention
- Multi-round Attention

MRC (Machine Reading Comprehension)

Performance-boosting Approaches for MRC Tasks

- Linguistic Embedding
- Multi-round Reasoning
- Reinforcement Learning
- Data Augmentation

MRC (Machine Reading Comprehension)

Transfer Learning in MRC Models

- Feature-based Transfer Learning
 - Context Vectors
 - Embedding from language Models
- Fine-tuning-based Transfer Learning
 - Generative Pre-training for Transformers
 - Bidirectional Encoder Representations from Trnasformers

MRC (Machine Reading Comprehension)

Open Issues

- Limitation of Given Context
- Robustness of MRC Systems
- Incorporation of External Knowledge
- Lack of Inference Ability
- Difficulty in Interpretation

MRC (Machine Reading Comprehension)

Conclusion

- definitions of MRC
- definitions of MRC tasks
- general architecture of neural MRC model

Appendix

dates	model	authors
2018/02	ELMo	Allen AI & UW
2018/05	GPT-1	OpenAI
2018/10	BERT	Google
2019/07	XLNet	CMU & Google Brain
2019/07	RoBERTa	FAIR
2019/09	ALBERT	Google & TTIC
2019/10	T5	Google

Thank You

contact : +82-10-5500-7977

Daumsoft Allab

kwmme797@gmail.com

<https://github.com/kwangjunechoi7>

