

SIRE: Separate Intra- and Inter-sentential Reasoning for Document-level Relation Extraction

ACL-IJCNLP Findings 2021

ING Lab 논문 세미나
2021-07-14

발표자: 현지웅

Document-level Relation Extraction

DocRED (ACL 2019) [paper] [dataset]

기존 Relation Extraction은 한 문장 내의 두 entity 간의 관계를 예측하는 task

Document-level Relation Extraction은 전체 문서 내의 두 entity 간의 관계를 예측하는 task로, 좀 더 어려운 문제

기존과 같이 한 문장만 보고 추측할 수 있는 문제도 있지만, 두 entity가 서로 떨어져 있는 경우 논리적인 추론이 필요함

평가지표: F1, Ign F1

* Ign F1은 train와 dev/test 사이의 relation facts에 대한 중복을 제외하고 구한 F1

ABBA Live

[1] ABBA Live is an album of live recordings by **Swedish** pop group ABBA, released by **Polar Music** in 1986. ... [6] The tracks were mostly taken from ABBA's concerts at **Wembley Arena** in **London** in November 1979. ... [13] It was remastered ...

Head: **Polar Music**

Tail: **Swedish**

relation: **country of origin**

evidence: [1]

Head: **Wembley Arena**

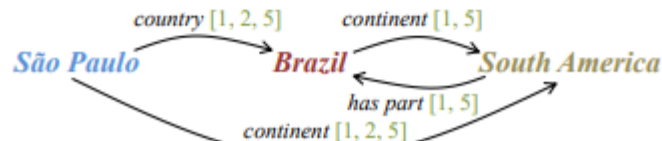
Tail: **London**

relation: **located in**

evidence: [6]

IBM Research-Brazil

[1] IBM Research-Brazil is one of twelve research laboratories comprising IBM Research, its first in **South America**. ... [2] It was established in June 2010, with locations in **São Paulo** and **Rio de Janeiro**. ... [5] In collaboration with **Brazil**'s government, it will help IBM ... [6] ... IBM has 4 priority areas in **Brazil** ...



Motivation

Relation types in DocRE

intra-sentential: 두 entity가 같은 문장 안에 등장하는 경우

inter-sentential: 그렇지 않은 경우

Previous work in DocRE

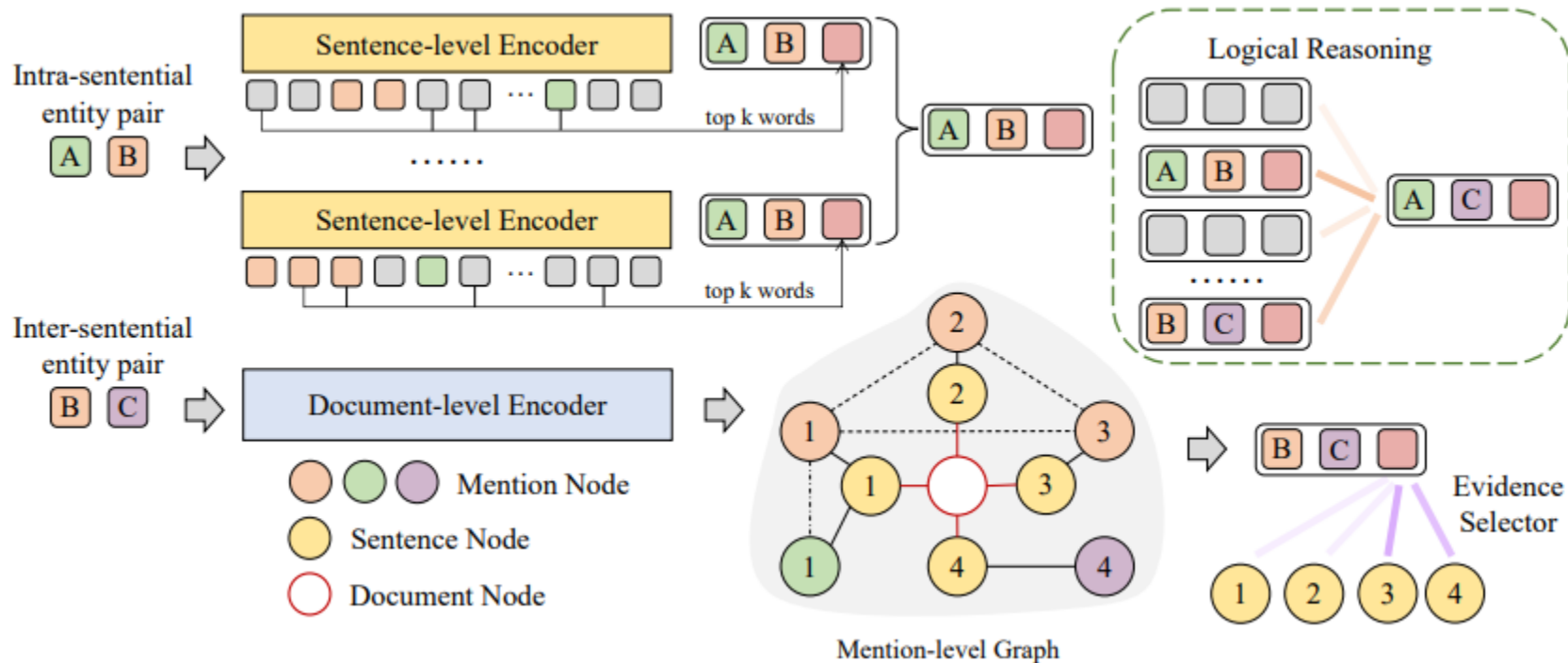
relation의 종류에 상관없이 모든 entity pair들에 대하여 가능한 relation을 예측하는 방식으로 접근

→ 언어학적 관점에서 intra-sentential과 inter-sentential은 다른 패턴으로 표현되어야 한다.

→ **SIRE** (Separate Intra- and Inter-sentential Reasoning)

SIRE

3개의 Module로 구성되어 있음 (Intra- and Inter-sentential Relation Representation Module, Logical Reasoning Module, Classification Module)



Intra-sentential Relation Representation Module

Encoding

각 문장(s_i)내 단어($w_j^{s_i}$)들을 시퀀스 벡터($g_j^{s_i}$)로 전환하는 단계

{word, entity type, co-reference} embedding을 concat하여
단어 벡터를 구성 (x)

$$\mathbf{x} = [E_w(w); E_t(t); E_c(c)]$$

단어 representation을 sentence-level encoder (f_{enc}^S)에 넣어
최종 시퀀스 벡터를 구성 ($g_j^{s_i}$)

$$[\mathbf{g}_1^{s_i}, \dots, \mathbf{g}_{n_i}^{s_i}] = f_{enc}^S([\mathbf{x}_1^{s_i}, \dots, \mathbf{x}_{n_i}^{s_i}])$$

sentence-level encoder: LSTM or BERT

Intra-sentential Relation Representation Module

Representing

각 entity pair $(e_{i,h}, e_{i,t})$ 에 대하여 intra-sentential한 관계를 표현하는 단계

head entity mention과 tail entity mention이 같은 문장에 등장하는 문장들의 집합 ($S_{co-occur}$)에 대하여, 각 문장별로 context representation을 구한다 $\{s_{i_1}, s_{i_2}, \dots, s_{i_C}\}$

* 저자진들은 context representation을 두 mention와 관련 높은 top K개의 단어들로 정의하였음

head entity mention과 tail entity mention을 query로 하여 문장 내 모든 단어들에 대한 relatedness(attention) score를 구하여 이를 기반으로 top K related word를 구한다

$$\mathbf{e}_{i,h}^{s_{ij}} = \frac{1}{t-s+1} \sum_{k=s}^t \mathbf{g}_k^{s_{ij}}$$
$$s_{i,k} = \sigma((W_{intra} \cdot [\mathbf{e}_{i,h}^{s_{ij}}; \mathbf{e}_{i,t}^{s_{ij}}])^T \cdot \mathbf{g}_k^{s_{ij}})$$
$$\alpha_{i,k} = \text{Softmax}(s_{i,k})$$

Intra-sentential Relation Representation Module

Representing

각 entity pair ($e_{i,h}, e_{i,t}$)에 대하여 intra-sentential한 관계를 표현하는 단계

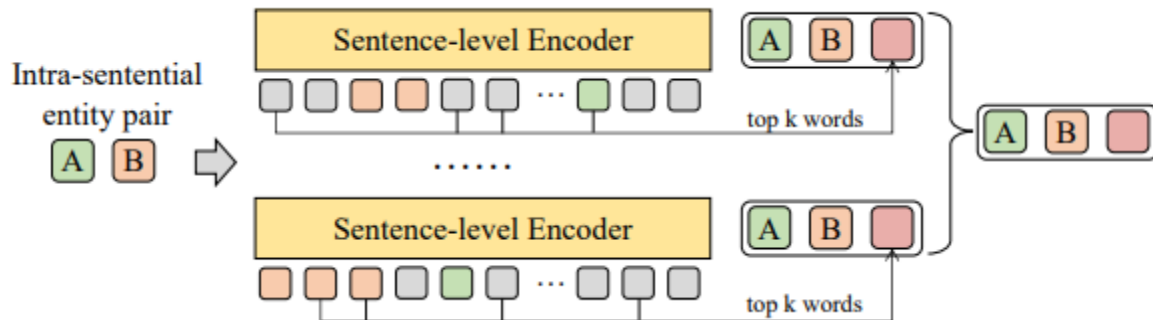
top K related word의 representation과 전체 weighted average representation을 더하여 context information ($c_i^{S_{ij}}$)을 구축

$$c_i^{S_{ij}} = \beta \cdot \frac{1}{K} \sum_{k \in \text{top}K(\alpha_{i,*})} g_k^{S_{ij}} + (1-\beta) \cdot \sum_t \alpha_{i,t} g_t^{S_{ij}}$$

* 이유: relatedness score를 구하기 위해 적용한 W_{intra} 에 대한 gradient를 구해야 학습이 가능하기 때문

최종 relational representation은 각 문장의 head, tail, context representation을 average

$$r_i = \frac{1}{C} \sum_{S_{ij} \in \mathcal{S}_{co-occur}} [e_{i,h}^{S_{ij}}; e_{i,t}^{S_{ij}}; c_i^{S_{ij}}]$$



Inter-sentential Relation Representation Module

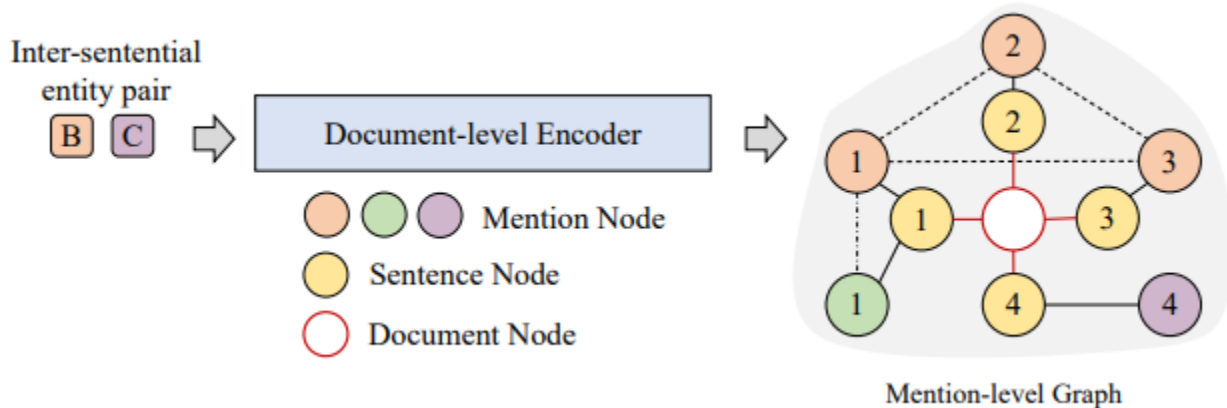
Encoding

각 문서(D)내 단어(w_j^D)들을 시퀀스 벡터(g_j^D)로 변환하는 단계

intra-와 동일한 방식으로 단어 representation을 구성

단어 representation을 document-level encoder (f_{enc}^D)에 넣어 시퀀스 벡터를 구성 (g_j^D)

mention-level graph (MG)를 적용하여 좀 더 문서 안의 상호 관계를 잘 표현할 수 있도록 함



Mention-level Graph

Mention-level Graph in GAIN (왼쪽)

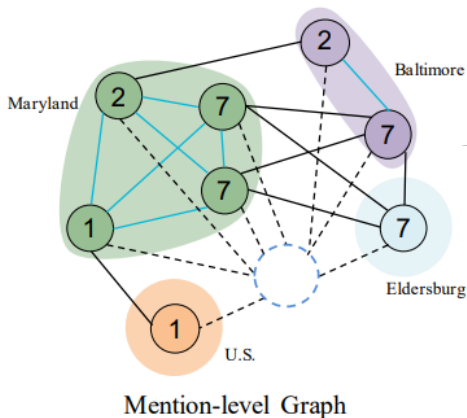
mention node와 document node로 구성

서로 다른 mention 끼리 document node를 pivot으로 하여 2단계만에 접근 가능

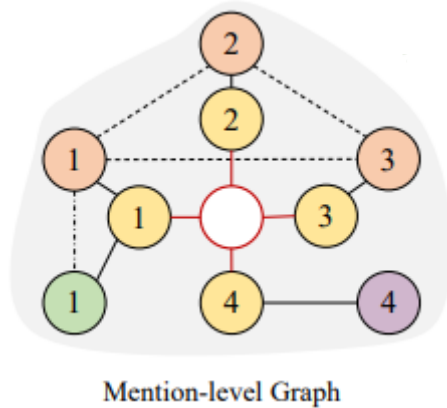
Mention-level Graph in SIRE (오른쪽)

GAIN의 MG는 local context information을 반영하고 있지 못함

따라서, sentence node와 그에 상응되는 edge를 추가하여 MG에 local information을 주입



- **Intra-Entity Edge:** 동일한 entity에 대한 mention끼리 연결
- **Inter-Entity Edge:** 동일 문장 내 서로 다른 entity에 대한 mention끼리 연결
- **Document Edge:** 문서 내 모든 mention과 document node와 연결



- **Sentence-Mention Edge:** 각 문장 내 mention와 문장과 연결
- **Sentence-Document Edge:** 각 문장과 document node와 연결
- 기존 Intra-, Inter-Entity Edge는 동일

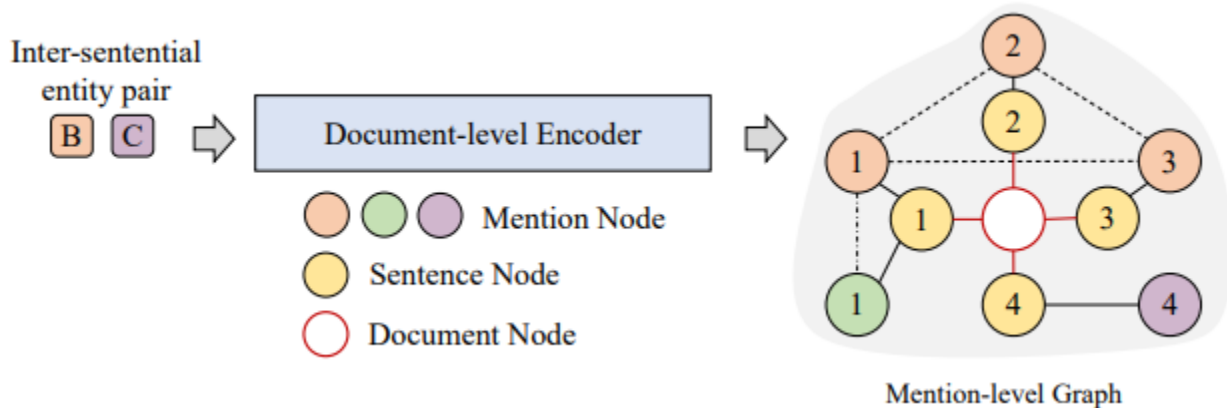
Inter-sentential Relation Representation Module

Encoding

각 문서(D)내 단어(w_j^D)들을 시퀀스 벡터(g_j^D)로 변환하는 단계

적용한 Mention Graph으로 부터 R-GCN 및 feature aggregation을 적용하면 각 node에 대한 representation을 뽑아낸다

$$\mathbf{h}_u^{(l+1)} = \text{ReLU} \left(\sum_{t \in \mathcal{T}} \sum_{v \in \mathcal{N}_u^t \cup \{u\}} \frac{1}{c_{u,t}} W_t^{(l)} \mathbf{h}_v^{(l)} \right) \quad \mathbf{h}_u^{(0)} = \frac{1}{t-s+1} \sum_{j=s}^t \mathbf{g}_j^D;$$
$$\mathbf{m}_u = \text{ReLU} (W_u \cdot [\mathbf{h}_u^{(0)}; \mathbf{h}_u^{(1)}; \dots; \mathbf{h}_u^{(N)}])$$



Inter-sentential Relation Representation Module

Representing

각 entity pair $(e_{i,h}, e_{i,t})$ 에 대하여 inter-sentential한 관계를 표현하는 단계

entity representation은 MG을 통해 구한 mention representation의 평균 $\mathbf{e}_i = \frac{1}{N} \sum_{j \in M(e_i)} \mathbf{m}_j$

전체 mention node에 대하여 attention을 적용하여 어느 mention이 추론의 evidence가 될 수 있을지 모델링 → context representation

$$P(\mathcal{S}_k | e_{i,h}, e_{i,t}) = \sigma(W_k \cdot [\mathbf{e}_{i,h}; \mathbf{e}_{i,t}; \mathbf{m}_{\mathcal{S}_k}])$$
$$\alpha_{i,k} = \frac{P(\mathcal{S}_k | e_{i,h}, e_{i,t})}{\sum_l P(\mathcal{S}_l | e_{i,h}, e_{i,t})} \quad \mathbf{c}_i = \sum_k \alpha_{i,k} \cdot \mathbf{m}_{\mathcal{S}_k}$$

최종 relation representation은 head, tail, context을 concat

$$\mathbf{r}_i = [\mathbf{e}_{i,h}; \mathbf{e}_{i,t}; \mathbf{c}_i]$$

Logical Reasoning Module

Logical Reasoning in Previous Work

각 entity pair 간들의 경로들을 단서로 하여 사용

문제: 모든 entity pair들이 연결되어 있지 않기도 하고 그래프 내에 올바른 추론 경로가 있다는 보장이 없음

Logical Reasoning in SIRE

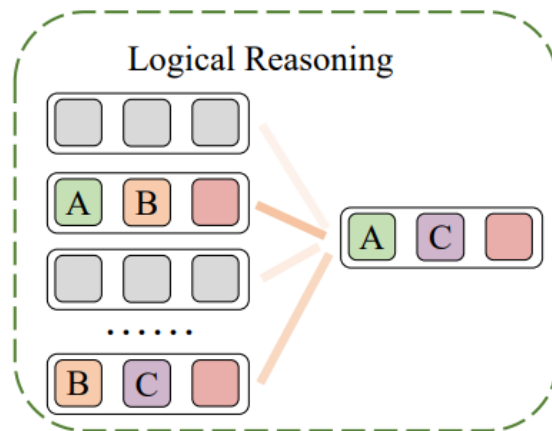
self-attention을 사용하여 logical reasoning을 모델링

한 entity pair (e_h, e_t)에 대하여 two-hop 형식의 logical reasoning chain $\{e_h \rightarrow e_k \rightarrow e_t\}$ 이 있다고 가정 (e_k 는 문서 내 존재하는 다른 entity)

$$\mathbf{r}_i^{new} = \sum_{\mathbf{r}_k \in \mathcal{R}_{att} \cup \{\mathbf{r}_i\}} \gamma_k \cdot \mathbf{r}_k$$

$$\gamma_k = \text{Softmax}((W_{att} \cdot \mathbf{r}_i)^T \cdot \mathbf{r}_k)$$

* \mathcal{R}_{att} 는 $(e_h, e_k), (e_k, e_t)$ 에 대한 relational representation



Classification Module

DocRE를 multi-label classification task로 formulate하여 학습

: two FC layer with sigmoid

$$P(r|e_{i,h}, e_{i,t}) = \text{sigmoid}(W_1 \sigma(W_2 \mathbf{r}_i + b_1) + b_2)$$

loss는 binary cross entropy

$$\begin{aligned} \mathcal{L}_{rel} = & - \sum_{\mathcal{D} \in \mathcal{C}} \sum_{h \neq t} \sum_{r_i \in \mathcal{R}} \mathbb{I}(r_i = 1) \log P(r_i | e_{i,h}, e_{i,t}) \\ & + \mathbb{I}(r_i = 0) \log (1 - P(r_i | e_{i,h}, e_{i,t})) \end{aligned}$$

Experiment

Dataset

DocRED: document-level 데이터셋 중 가장 큰 규모

CDR: Chemical-Disease Reactions 데이터셋

GDA: Gene-Disease-Associations 데이터셋

Experiment Settings

SIRE-GloVe: word embedding (GloVe 100d) + encoder (BiLSTM 512d)

SIRE-BERT: word embedding (BERT-base) + encoder (BERT-base)

*CDR/GDA는 BioBERT-Base v1.1 사용

Baselines

BiLSTM, BERT, HIN-GloVe, HIN-BERT, CorefBERT, LSR-GloVe, LSR-BERT, GLRE-BERT, GAIN-BERT, BRAN

Experiment

Model	Dev				Test	
	Ign F1	F1	Intra-F1	Inter-F1	Ign F1	F1
BiLSTM (Yao et al., 2019b)	48.87	50.94	57.05	43.49	48.78	51.06
HIN-GloVe (Tang et al., 2020)	51.06	52.95	-	-	51.15	53.30
LSR-GloVe (Nan et al., 2020)	48.82	55.17	60.83	48.35	52.15	54.18
GAIN-GloVe (Zeng et al., 2020)	53.05	55.29	61.67	48.77	52.66	55.08
SIRE-GloVe	54.10	55.91	62.94	48.97	54.04	55.96
-LR Module	53.73	55.58	62.77	47.87	53.75	55.55
-context	52.57	54.41	61.66	46.92	52.33	54.15
-inter4intra	52.23	54.26	60.81	48.36	51.77	53.30
BERT (Wang et al., 2019a)	-	54.16	61.61	47.15	-	53.20
BERT-Two-Step (Wang et al., 2019a)	-	54.42	61.80	47.28	-	53.92
HIN-BERT (Tang et al., 2020)	54.29	56.31	-	-	53.70	55.60
CorefBERT (Ye et al., 2020)	55.32	57.51	-	-	54.54	56.96
GLRE-BERT (Wang et al., 2020)	-	-	-	-	55.40	57.40
LSR-BERT (Nan et al., 2020)	52.43	59.00	65.26	52.05	56.97	59.05
GAIN-BERT (Zeng et al., 2020)	59.14	61.22	67.10	53.90	59.00	61.24
SIRE-BERT	59.82	61.60	68.07	54.01	60.18	62.05

Model	CDR	GDA
BRAN (Verga et al., 2018)	62.1	-
EoG (Wang et al., 2020)	63.6	81.5
LSR (Nan et al., 2020)	64.8	82.2
GLRE-BioBERT (Wang et al., 2020)	68.5	-
SIRE-BioBERT	70.8	84.7

Ablation Study

SIRE-GloVe에 대하여 실험 진행

Logical Reasoning Module의 효과를 검증하기 위한 실험

logical reasoning module 제거 후 F1 0.41 하락

context representation의 효과를 검증하기 위한 실험

context 제거 후 F1 1.81 하락

intra-와 inter-의 구분 필요성을 검증하기 위한 실험

intra-sentential entity pairs에 대해서도 inter-sentential module을 적용했을 때,
F1 2.66, Intra-F1 2.13 하락

Model	Dev				Test	
	Ign F1	F1	Intra-F1	Inter-F1	Ign F1	F1
SIRE-GloVe	54.10	55.91	62.94	48.97	54.04	55.96
-LR Module	53.73	55.58	62.77	47.87	53.75	55.55
-context	52.57	54.41	61.66	46.92	52.33	54.15
-inter4intra	52.23	54.26	60.81	48.36	51.77	53.30

Reasoning Performance

Reasoning Ability

dev set 안에 two-hop을 통해 예측할 수 있는 relation에 대해서만 F1 score를 평가

기존 모델 배이 Infer-F1 점수가 상당히 높고, Logical Reasoning module을 제외했을 때 그만큼 성능 감소

Model	Infer-F1	P	R
BiLSTM	38.73	31.60	50.01
GAIN-GloVe	40.82	32.76	54.14
SIRE-GloVe	42.72	34.83	55.22
- <i>LR Module</i>	39.18	31.97	50.59

Case Study

Case Study on Intra- and Inter-sentential relation instances

Intra의 경우 잘 예측함을 확인할 수 있음

Inter의 경우, 각 문장 별 attention score를 보면 [1], [2] 문장을 모델이 supporting evidence로 판단 (점수 0.87, 0.66)

Sao Paulo와 South America 사이의 logical reasoning chain에서 Sao Paul - Brazil - South America 로 올바른 reasoning 수행 [(Sao Paulo, Brazil) : 0.32, (Brazil, South America): 0.45)

Type	Examples											
Intra-sentential relation instances	<p>"<i>Your Disco Needs You</i>" is a song performed by Australian recording artist and songwriter <i>Kylie Minogue</i>, taken from her seventh studio album <i>Light Years</i> (2000). Relation: performer</p> <p><i>Lark Force</i> was an <i>Australian Army</i> formation established in <i>March 1941</i> during <i>World War II</i> for service in <i>New Britain</i> and <i>New Ireland</i>. Relation: inception</p> <p><i>Lake Hiawatha</i> is one of the few lakes through which <i>Minnehaha Creek</i> flows, and the last one before it reaches <i>Minnehaha Falls</i> and then the <i>Mississippi River</i>. Relation: mouth of the watercourse</p>											
	Inter-sentential relation instances	<p>[1] (0.87) <i>IBM Research–Brazil</i> is one of twelve research laboratories comprising <i>IBM Research</i>, its first in <i>South America</i>. [2] (0.66) It was established in <i>June 2010</i>, with locations in <i>São Paulo</i> and <i>Rio de Janeiro</i>. [3] (0.01) Research focuses on Industrial Technology and Science, ... [4] (0.04) The new lab, <i>IBM</i>'s ninth ... [5] (0.38) In collaboration with <i>Brazil</i>'s government, it will help <i>IBM</i> to develop ... Relation: continent</p> <p>Logical reasoning attention weight:</p> <table><tr><td>(São Paulo, Brazil)</td><td>0.32</td><td>(São Paulo, June 2010)</td><td>0.03</td><td>...</td></tr><tr><td>(Brazil, South America)</td><td>0.45</td><td>(June 2010, South America)</td><td>0.02</td><td>...</td></tr></table>	(São Paulo, Brazil)	0.32	(São Paulo, June 2010)	0.03	...	(Brazil, South America)	0.45	(June 2010, South America)	0.02	...
	(São Paulo, Brazil)	0.32	(São Paulo, June 2010)	0.03	...							
(Brazil, South America)	0.45	(June 2010, South America)	0.02	...								

Conclusion

- 저자진의 이전 연구 또한 SOTA였으나 다시 SOTA를 달성한 연구
- 자신의 기존 연구를 진행하면서 생각한 문제점을 해결하고자 하는 모습이 보임
- Graph 기반이라 RoBERTa-large와 같은 큰 encoder를 사용하는 것이 어려운 것으로 추정? (이 때의 성능이 궁금)