

Conditional Generation of Temporally-ordered Event Sequences

2021-08-05

성창민

Introduction

- Propose

- 하나의 모델로 temporal event ordering과 event infilling 두 가지 tasks 수행
- 시나리오에서 나타나지 않은 사건 추론

- Goal

- 폭 넓게 temporal event knowledge를 습득
- 넓은 범위의 추론

- Model

- TemporalBART(BART-based)
- 기존의 모델들은 일반적인 사건의 시간 지식보다는 대화에서의 신호들을 학습

Introduction

- Approach

- Distributed representations of schemas를 사용하여 추론

- Motivation

- Schema learning
 - 사건의 전, 후 그리고 사이에 어떤 사건이 발생하는지 이해하는 text generation tasks
 - 원인, 영향 등에 대한 answering questions

Introduction

Scrambled input events

I opened a present
She bought the present

Temporal
BART

Complete ordered event sequence

She bought the present
She gave me the present
I opened a present

- TemporalBART

- 인코더는 원본 문장들 중 일부를 추출한 후 순서를 섞은 문장들을 입력 받음
- 디코더는 사라진 문장과 순서가 복원된 문장들을 출력함

Introduction

- Data

- EventsNarratives 데이터셋에서 필요한 부분을 자동으로 추출
- 넓은 도메인 범위의 데이터

- Evaluation

- Zero-shot 방법으로 out-of domain test 수행
- CaTeRS와 MCTaco 데이터셋 사용
- BERT-based pairwise 모델과 BERT-based pointer network 모델로 비교
- HAQAE 모델, GPT-2 모델과 Infilling GPT-2 모델로 비교

Background and Related Work

- Temporal Event Ordering
 - Temporal relation extraction
 - 이후 연구는 주로 pairwise relations에 초점을 둠
 - 이 논문에서는 Sequence generation problem으로서의 **사건 순서**에 중점을 두고자 함
- Schema Induction
 - 통계적 방법 & 사건 분포 학습 등의 다양한 연구
 - 대화의 순서를 학습하도록 빠지기 쉬움

Method

- Task Formulation and Model

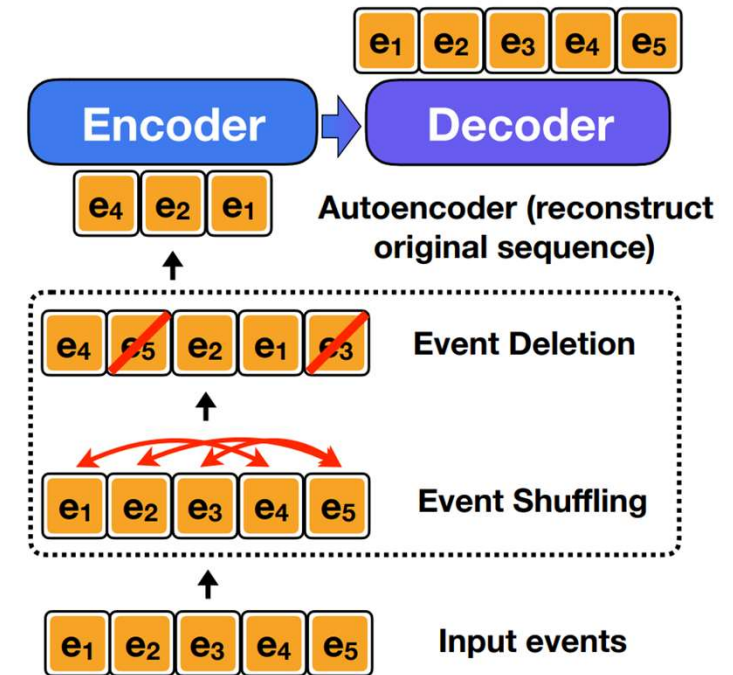
- $P(y|x)$ over temporal event sequences $y = \{e_1, \dots, e_l\}$, y 는 순서대로 이루어진 사건 집합
- Input $x = \{e_1, \dots, e_m\}$, x 는 순서대로 이루어지지 않은 y 의 부분 집합
- 사건의 시간적 지식을 학습하는 것을 의도

- Event Shuffling

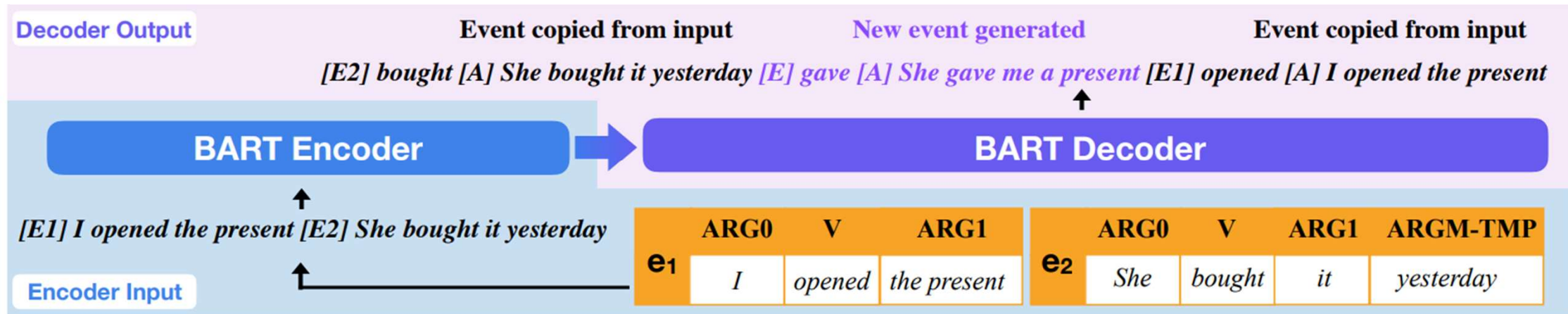
- y 를 x 로 만들기 위해 랜덤하게 섞음

- Event Deletion

- 일정 확률 p 로 y 에서 각각의 e 를 제거하여 x 를 생성
- $\log P(y|x)$ 를 최대화하도록 학습



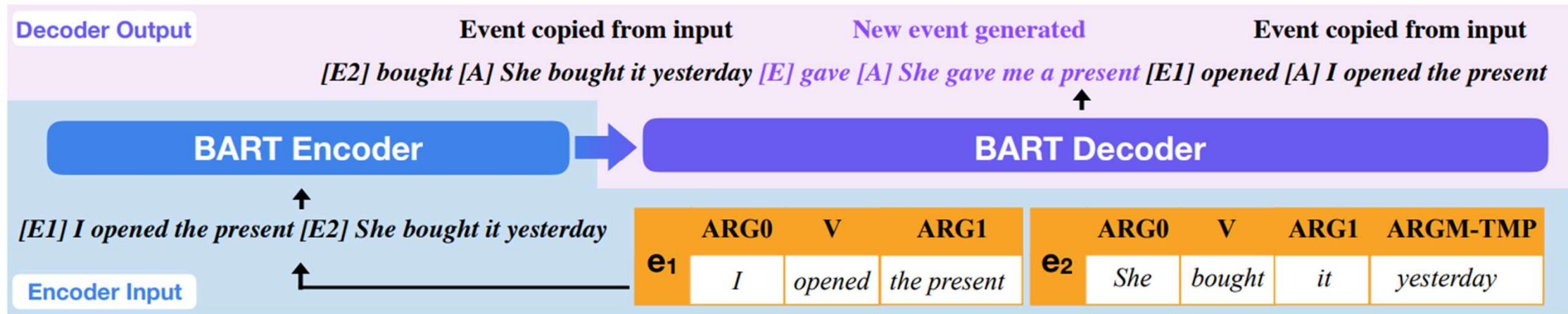
Method



• Model Architecture

- BART의 pretrained weights로 모델 initialize
- 각 사건의 predicate와 arguments를 concatenation하여 e를 표현
- 명사구 arguments를 보존하여 인코드

Method



• TemporalBART

- 사건 e 를 표현한 후 $[E]$ 의 special token을 concatenate하여 입력
- Predicate를 기반으로 순서 예측을 의도 및 확인하기 위해 $[E] v_{e_j} [A]$ 를 출력

• TemporalBART-indexed

- TemporalBART의 $[E]$ 대신 $[E_i]$ 을 사용하여 순서 학습에 도움을 줌

Method

- Training Data Collection
 - 대화 순서가 아닌 사건 순서의 데이터가 필요함
 - 일찍 쓰여진 사건이 시간상으로 나중에 일어난 경우가 많음
- Corpus
 - EventsNarratives corpus 사용
 - 뉴스, 소설과 블로그 domain으로부터 데이터를 수집(소설에 가장 초점을 맞춤)
- Extracting Temporal Event Sequences
 - AllenNLP의 SRL 모델을 사용하여 verbs(events)와 arguments를 추출
 - 다른 문장에 있는 events를 연결하여 sequences 만듦
 - 사건이 연결되어 있음을 보장하기 위해 common entity가 있는 events로만 구성

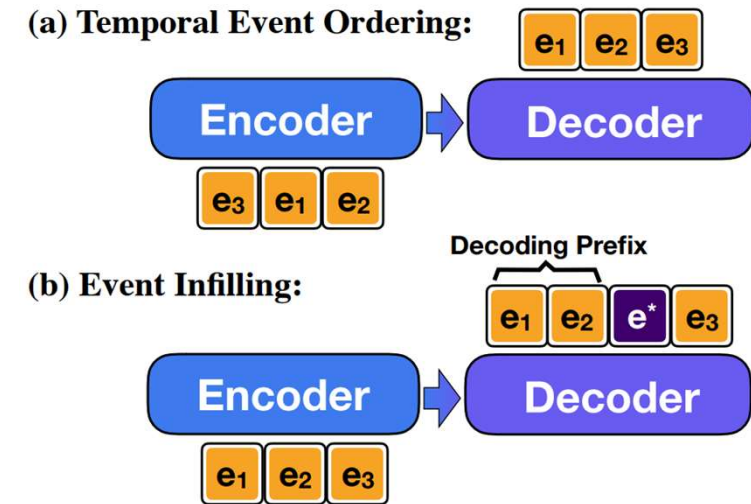
Target Task Formulation

- Temporal Event Ordering

- 무작위 순서의 $\{e_i\}$ 를 입력 받아 시간 순서의 $\{e_i\}$ 를 출력

- Event Infilling

- Pre-selected insertion positions에 들어갈 events를 생성
- 입력 $x = \{e_i\}$ 를 받아 i^* 순서에 event를 생성
- 디코더의 prefix로 $x_{\text{prefix}} = \{e_i | i < i^*\}$ 를 사용하여 e^* 를 출력

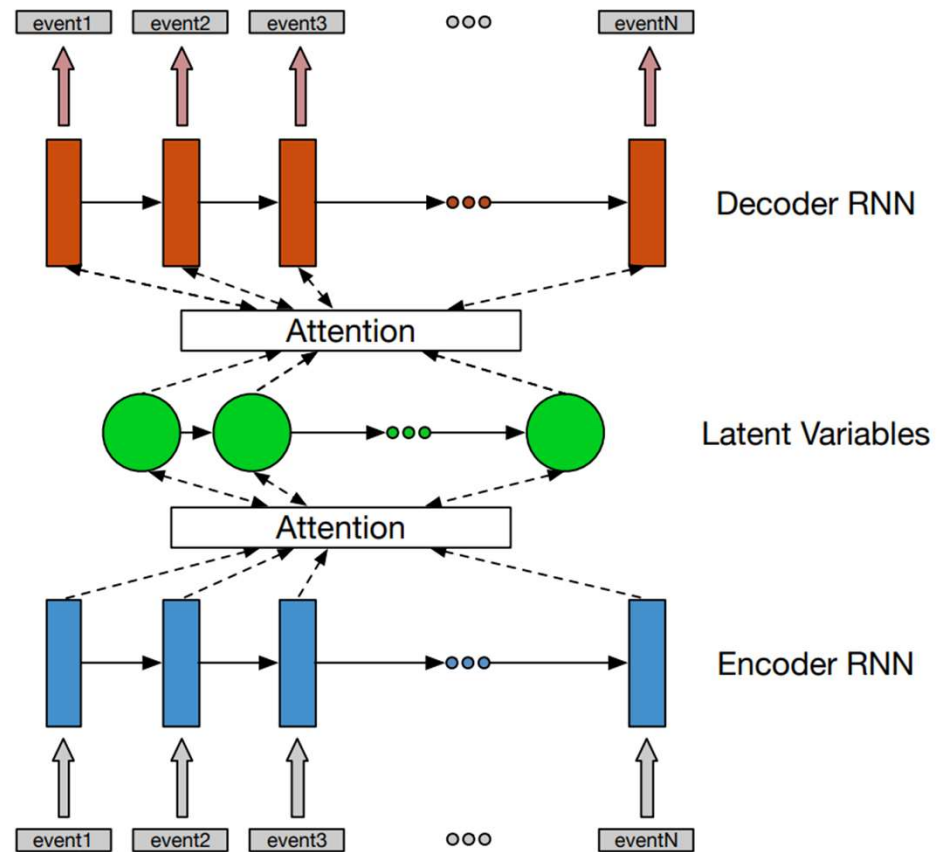


Target Task Formulation

- Baselines: Temporal Event Ordering
 - BERT-based Pairwise Model + SSVM(Structured Support Vector Machine)
 - Event ordering as a pairwise classification problem
 - 출력에서 e_i 가 e_j 에 앞서는 지의 pairwise scores를 계산
 - BERT-based Pointer Network
 - BERT-based Pairwise Model + SSVM 모델에서 U_{p_i} 벡터를 추출
 - U 는 BERT의 encoded matrix, p_i 는 e_i 의 첫 번째 토큰의 위치, h_t 는 디코더의 hidden state
 - LSTM-based pointer network 모델로 ordering probability 계산
 - $$P^{\text{seq}}(\mathbf{y} \mid \mathbf{x}) = \prod_j P(j \mid \mathbf{h}_1, \dots, \mathbf{U}_{p_1}, \dots)$$

Target Task Formulation

- Baselines: Event Infilling
 - HAQAE
 - Hierarchical Quantized Autoencoder
 - Event-level seq2seq autoencoder
 - N+1개의 event가 출력될 것으로 추정



Target Task Formulation

- Baselines: Event Infilling
 - GPT-2
 - Only perform uni-directional generation
 - 디코더의 prefix로 $x_{\text{prefix}} = \{e_i | i < i^*\}$ 를 사용
 - 출력된 값을 e^* 로 해석
 - BART-large 모델과 크기가 비슷한 GPT2-medium pre-trained 모델을 사용
 - Infilling GPT-2
 - Infilling tasks에 맞도록 GPT-2의 디코더의 prefix를 일부 수정
 - 디코더의 prefix로 $\{\text{Repr}(e_i) | i \geq i^*\}$, $\langle \text{SEP} \rangle$ 그리고 $\{\text{Repr}(e_i) | i < i^*\}$ 를 concatenation 한 것을 사용
 - $\langle \text{SEP} \rangle$ 토큰이 사건의 전후를 구분하는데 도움을 줌

Evaluation

- Experimental Setup

- GPT-2를 제외한 모든 모델은 구축한 데이터셋으로 학습
- 100,000개의 데이터를 2 permutations로 하여 200,000개의 데이터를 학습
- Zero-shot 방법으로 out-of-domain 데이터셋을 사용

Evaluation

• Temporal Event Ordering

• Datasets

- CaTeRS와 MCTaco 사용
- CaTeRS
 - ROCStories에서 샘플링된 short story
 - 추출된 데이터의 60%가 3개 이상의 event로 구성
- MCTaco
 - Multiple-choice QA 데이터셋
 - Event ordering 추론을 요구하는 데이터를 추출
 - e^q 가 e^a 보다 전인지 후인지 판단
 - CaTeRS보다 다양한 domain

Context:

In Colombia, the drug-financed guerrillas trying to seize the country and destroy democracy include M-19, which Castro has clearly backed.

Question:

What would the guerrillas do if able to seize the country ?

Candidate Answer:

they would destroy the democracy

Extracted Event Sequence:

e1: drug financed guerrillas

e2: the drug - financed guerrillas trying to seize the country and destroy democracy

e3: the drug - financed guerrillas seize the country

e4: the drug - financed guerrillas destroy democracy

e5: In Colombia the drug - financed guerrillas trying to seize the country and destroy democracy include M-19 , which Castro has clearly backed

e6: M-19 which Castro clearly backed

e7: they would destroy the democracy

Gold Label:
 e^a AFTER e^q



Evaluation

- Temporal Event Ordering

- Results on CaTeRS

Architecture	All Pairwise Acc.	Length ≥ 3 Pairwise Acc.
Random	50.4	50.2
Pairwise+SSVM	65.7	62.3
BERT-based PN	54.1	52.3
TemporalBART	77.1	74.7
TemporalBART-indexed	79.7	78.0

- Results on MCTaco

Architecture	Acc.	Macro F1
Majority	90.6	47.5
Pairwise+SSVM	67.2	47.0
BERT-based PN	54.7	42.7
TemporalBART	63.9	50.1
TemporalBART-indexed	74.9	55.1

Evaluation

- Ordering Unseen Events

- $e_1, e_2, \dots, e^*, \dots, e_N$ 이 있을 때, e^* 를 제거한 것을 입력
- e^* 를 제거한 부분을 맞추는 실험(e_1 이 정답일 때 e 이 EM, e 가 Top2 EM)

Architecture	All		Length ≥ 3	
	EM	Top2 EM	EM	Top2 EM
Random	34.1	69.5	23.7	48.7
HAQAE	37.1	71.9	28.7	53.2
GPT-2	35.2	68.4	22.6	48.2
Infilling GPT-2	38.8	73.5	26.3	55.4
TemporalBART	57.7	83.3	48.2	70.6
TemporalBART-indexed	58.4	87.4	50.9	77.4
- event deletion	42.4	73.0	29.8	53.8

- e_1 이 정답일 때 e_1 이 30%, e_2 가 50%, e_3 가 20%인 경우 Top2 EM
- Event deletion이 많은 영향을 미침

Evaluation

- Event Infilling

- i *번째 event를 제거한 것을 입력
- i *번째에 어떠한 event를 생성하는 지를 실험

Architecture	Coherence	Temporality
GPT-2	1.37	0.57
Infilling GPT-2	1.50	0.87
TemporalBART	1.43	1.10
TemporalBART-indexed	1.50	1.03

GPT-2: You can buy a \$25 Apple Watch with the watch face

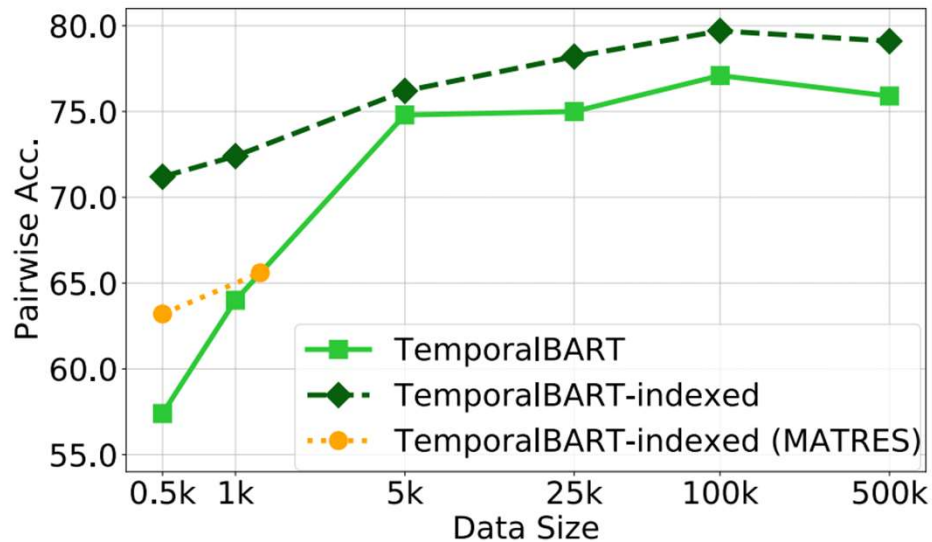
Infilling GPT-2: He started with a rough - looking tree

TemporalBART: After breakfast Mike picked a good piece of twine

[INSERTED EVENT] **e2:** Mike tried to make a tree **e3:** He painted over the bad tree with design of his own creation

Evaluation

- The Effectiveness of Narrative Data



- 데이터 크기가 증가할수록 성능 향상
- MATRES 데이터셋으로 학습하면 성능 하락
- CaTeRS가 temporal knowledge 학습 평가하기 좋은 데이터셋