Modeling Event-Pair Relations in External Knowledge Graphs for Script Reasoning

Yucheng Zhou[1]*, **Xiubo Geng**[2]†, **Tao Shen**[3], **Jian Pei**[4], **Wenqiang Zhang**[1], **Daxin Jiang**[2]†
[1]Fudan University, Shanghai, China
[2]Microsoft, Beijing, China
[3]Australian AI Institute, School of CS, FEIT, University of Technology Sydney
[4]Simon Fraser University, Burnaby, Canada
{yczhou18,wqzhang}@fudan.edu.cn, {xigeng,djiang}@microsoft.com
tao.shen@student.uts.edu.au, jpei@cs.sfu.ca

Abstract

Script reasoning infers subsequent events from a given event chain, which involves the ability to understand relations between events. A human-labeled script reasoning dataset is usually of small size with limited event relations, which highlights the necessity to leverage external eventuality knowledge graphs (KG) consisting of numerous triple facts to describe the inferential relation between events. Existing methods adopt a *retrieval and integration* paradigm to focus merely on the graph triples that have event overlap with a script, but ignore much more supportive triples in the KG with similar inferential patterns, leading to under-exploiting. To fully exploit the KG, we propose a knowledge model to learn the inferential relations between events from the whole eventuality KG and then support downstream models by directly capturing the relation between events in a script. We further present a neural script adapter to extend the knowledge model for inferring the associated relations between an event chain and a subsequent event candidate. We evaluate the proposed approach on a popular multi-choice narrative cloze task for script reasoning and achieve new state-of-the-art accuracy, compared with baselines either incorporating external KG or not.

1 Introduction

Script reasoning (Chambers and Jurafsky, 2008; Li et al., 2018; Lv et al., 2020b) aims at determining the subsequent event or plausible ending for an event chain in a script. For example, a tourism script consist of [“*Emily took a plane*”, “*Emily arrived at Oahu*”, “*Emily went to Waimea Bay*”], and the subsequent event is more likely to be “*Emily surfed*” than “*Emily skied*”. Script reasoning has attracted more interest in the natural language processing (NLP) community since it plays essential roles in many real-world applications like storytelling (Swanson and Gordon, 2008).

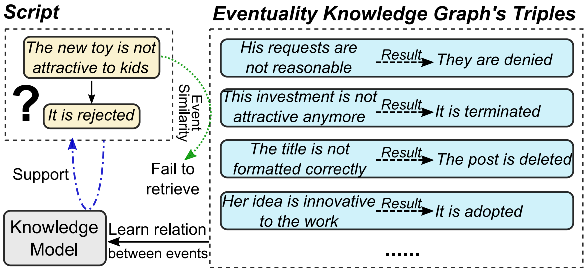
Script
Eventuality Knowledge Graph's Triples
The new toy is not attractive to kids
?
It is rejected
Event Similarity
Support
Fail to retrieve
Knowledge Model
Learn relation between events
His requests are not reasonable
Result
They are denied
This investment is not attractive anymore
Result
It is terminated
The title is not formatted correctly
Result
The post is deleted
Her idea is innovative to the work
Result
It is adopted
......

Figure 1: Comparison of “*retrieval and integration*” paradigm (green dot line) with ours (blue dash-dot line). Although there is no semantic overlap between the precedent event in the script and the events in the KG, which leads to failed retrieval, our approach still provides supportive evidence by exploiting similar inferential relation patterns.

Understanding and inferring the correlation between two events are critical for script reasoning. Taking the tourism script as an example, the key to decide the subsequent event is inferring that “*A person goes to a beach*” is more correlated to “*The person surfs*” than “*The person skies*”. An immediate idea is to learn event relations from some well-labeled training datasets. Unfortunately, due to labor-intensive labeling, high-quality training data for script reasoning is usually small, from which it is impractical to learn rich relations for large scale commercial applications. Therefore, it is necessary to leverage external knowledge that implies relations between events.

Recently, Lv et al. (2020b) propose to leverage a large-scale eventuality knowledge graph (KG), ASER (Zhang et al., 2020), for script reasoning via adopting the “*retrieval and integration*” paradigm. Given an event chain, this paradigm first retrieves relevant triple facts from the eventuality KG and then integrates them into a script reasoning model.

Although such a paradigm is proven effective in entity-centric tasks (Zhang et al., 2019; Liu et al., 2020), it is not competent in event-centric script

[*] Work is done during internship at Microsoft.
[†] Corresponding author.

Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 4586–4596
August 1–6, 2021. ©2021 Association for Computational Linguistics

reasoning. The reason is that, the retrieval is based on lexical or semantic matching between an event from a script and each event node in the KG. For example, in Figure 1, to determine whether the precedent event "*The new toy is not attractive to kids*" will result in a subsequent event "*It is rejected*", this paradigm will try to retrieve graph triples with the event nodes talking about "*toy is not attractive*", etc., which is very likely to fail if the KG contains few related events. Namely, it dramatically narrows the focus to the graph triples merely with exact event matching, so it cannot fully leverage the external eventuality KG.

However, script reasoning can benefit from leveraging *event pairs* in KG with *similar relation patterns*, rather than the only triples in KG with *similar events*. In Figure 1, although events in the four graph triples have no semantic overlap with the precedent event "*The new toy is not attractive to kids*", all the triples can represent the relation that *if some attribute of an object is judged negatively, it might be rejected, otherwise being accepted*, which still provide strong supportive evidence between "*The new toy is not ...*" and "*It is rejected*". Therefore, script reasoning can benefit from the event pairs with similar inferential relation patterns, beyond the textual contents of the events.

Motivated by this, in this work, we propose a novel paradigm to integrate external knowledge for script reasoning by directly modeling the relation between events from a KG and thus support script reasoning in light of similar relation patterns. In particular, we first propose a *discriminative knowledge model* trained on the graph triples in an external eventuality KG. Taking each event pair in the triples as input, the knowledge model learns to predict whether two events in the pair are associated and what is the inferential relation between them. After being trained, the knowledge model can directly capture associated and inferential relations between precedent and subsequent events in a script. And the relations between events will be represented in latent space, which can be further integrated into any event-centric neural model.

Furthermore, as script reasoning requires to associate between a sequence of precedent events (i.e., an event chain) and a plausible subsequent event, we propose a neural *script adapter*, based on a chain-dependent attention module, for extending the trained knowledge model from event to script level. This leads to a *script-adaptive knowledge model* that directly represents inferential information between an event chain and a subsequent event candidate as a latent embedding. Lastly, this embedding, coupled with deep text representation from a script-text contextualizing encoder, is used to derive the plausibility score of the candidate.

We conduct empirical studies on a popular task of script reasoning, i.e., multi-choice narrative cloze (Li et al., 2018). Our approach outperforms strong competitors and achieves a new state-of-the-art accuracy, verifying the effectiveness of the script-adaptive knowledge model when integrating inferential relations into script reasoning.

2 Preliminary

This section begins with a formal task definition of script reasoning, followed by introductions to eventuality KGs and pre-trained language models.

Task Definition. Script reasoning is usually formulated as a multi-choice narrative cloze (MCNC) problem: given an event chain $\boldsymbol{e} = [e_1, \ldots, e_n]$, it aims to select the most plausible subsequent event from a set of candidates $\mathcal{E}^{(c)} = \{e_1^{(c)}, \ldots e_m^{(c)}\}$, where each event e consists of a sequence of words $\boldsymbol{w}^e = [w_1^e, w_2^e, \ldots]$, n denotes the length of event chain $\boldsymbol{e}$, and m denotes the number of candidates $\mathcal{E}^{(c)}$. A script reasoning model is asked to produce relatedness score between the event chain and each candidate event so that

$$\hat{e} = \arg\max_{e_j} P(\boldsymbol{e}, e_j; \theta), \ \forall e_j \in \mathcal{E}^{(c)}, \qquad (1)$$

where $P(\cdot; \theta)$ denotes a θ-parameterized script reasoning model, and $\hat{e}$ denotes the predicted event.

Eventuality Knowledge Graph. In contrast to *canonical KGs* with entity-centric factoid triples, an eventuality KG, $\mathcal{G}$, typically consists of a set of event-centric triples $(e^{(h)}, r, e^{(t)})$ to describe inferential or co-occurrent relation between events. It represents each event e as free-form text, while well defines a closed-set $\mathcal{R}$ of relations so that $\forall r \in \mathcal{R}$.

3 Methodology

In this section, we will elaborate on our approach for multi-choice narrative cloze (MCNC) task in script reasoning. As shown in Figure 2, we first propose a discriminative knowledge model learning facts from eventuality graph (§3.1), followed by a novel neural adapter upgrading the knowledge model into script level (§3.2). Lastly, as in Figure 3,

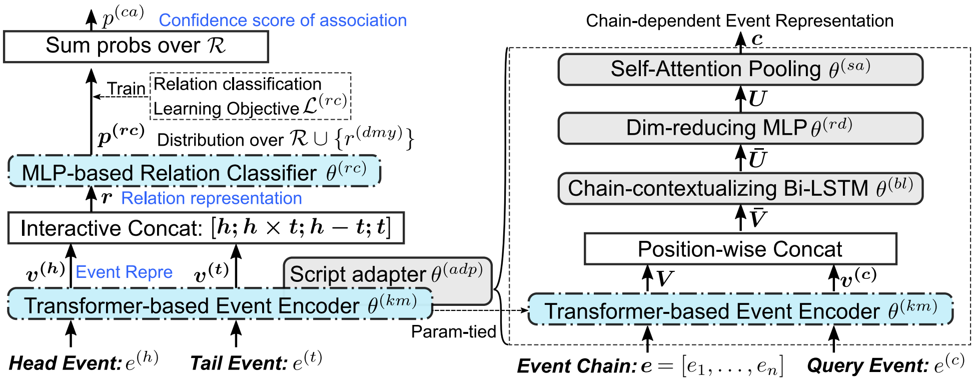
$p^{(ca)}$ Confidence score of association
Sum probs over $\mathcal{R}$
Train
Relation classification Learning Objective $\mathcal{L}^{(rc)}$
$\boldsymbol{p}^{(rc)}$ Distribution over $\mathcal{R} \cup \{r^{(dmy)}\}$
MLP-based Relation Classifier $\theta^{(rc)}$
$\boldsymbol{r}$ Relation representation
Interactive Concat: $[\boldsymbol{h}; \boldsymbol{h} \times \boldsymbol{t}; \boldsymbol{h} - \boldsymbol{t}; \boldsymbol{t}]$
$\boldsymbol{v}^{(h)}$ Event Repre $\boldsymbol{v}^{(t)}$
Script adapter $\theta^{(adp)}$
Transformer-based Event Encoder $\theta^{(km)}$
Param-tied
Head Event: $e^{(h)}$
Tail Event: $e^{(t)}$
Chain-dependent Event Representation
$\boldsymbol{c}$
Self-Attention Pooling $\theta^{(sa)}$
$\boldsymbol{U}$
Dim-reducing MLP $\theta^{(rd)}$
$\bar{\boldsymbol{U}}$
Chain-contextualizing Bi-LSTM $\theta^{(bl)}$
$\bar{\boldsymbol{V}}$
Position-wise Concat
$\boldsymbol{V}$
$\boldsymbol{v}^{(c)}$
Transformer-based Event Encoder $\theta^{(km)}$
Event Chain: $\boldsymbol{e} = [e_1, \ldots, e_n]$
Query Event: $e^{(c)}$

Figure 2: Our discriminative knowledge model (left), and its script adapter (right) for multi-choice narrative cloze (MCNC). Dash-dot blue rounded rectangles denote parameters optimized towards the objective of the knowledge model, whereas solid blue rounded rectangles denote script adapter's parameters that will be optimized towards the objective of MCNC.

we present a representation learning framework to solve the MCNC task (§3.3).

3.1 Discriminative Knowledge Model

To avoid challenging event grounding and satisfy coverage necessity, neural *knowledge models* (Bosselut et al., 2019b; Hwang et al., 2020) are proposed to memorize eventuality facts from a KG to its parameters during training. They are built upon a pre-trained *generative* Transformer (e.g., GPT (Radford et al., 2018)) and fine-tuned on triple facts from an eventuality KG via generative objectives of event-based *link prediction*.

However, such generative knowledge models are not perfectly compatible when capturing event-pair relation facts since they focus more on inferring tail events given a head event and an inferential relation. This is consistent with the goal of link prediction for KG completion. Consequently, if they try to model the inferential relations between events, they have to generate all possible triples for each event by traversing all relations and enlarging beam-search size (Bosselut et al., 2019a). And the generated triple must be re-encoded into latent space for the integration (Lv et al., 2020b), not to mention generative models not always reliable.

Therefore, we present a discriminative objective based on *relation classification* for knowledge model learning to directly capture such inferential information in latent space. Formally, given a triple $(e^{(h)}, r, e^{(t)}) \in \mathcal{G}$, we separately pass head event $e^{(h)}$ and tail event $e^{(h)}$, into a text encoder to generate event-level contextualized representations. Following Devlin et al. (2019), we first concatenate the natural language text $\boldsymbol{w}^e$ of each event e with special tokens:

$$\tilde{\boldsymbol{w}}^e = (\texttt{[CLS]}, \boldsymbol{w}^e, \texttt{[SEP]}), \forall e \in \{e^{(h)}, e^{(t)}\}, \quad (2)$$

where the special tokens could vary with different pre-trained models. Then, we feed the concatenated text $\tilde{\boldsymbol{w}}^e$ into a Transformer encoder, followed by a pooling layer, i.e.,

$$\boldsymbol{H}^e = \text{Transformer}(\tilde{\boldsymbol{w}}^e; \theta^{(km)}) \in \mathbb{R}^{d \times N}, \quad (3)$$

$$\boldsymbol{v} = \text{Pool}(\boldsymbol{H}^e) \in \mathbb{R}^d, \ \forall e \in \{e^{(h)}, e^{(t)}\}, \quad (4)$$

where $\boldsymbol{v}$ denotes the resulting event representation, $\text{Transformer}(\cdot; \theta)$ stands for pre-trained bidirectional Transformer encoder (e.g., BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019)) to produce deep contextualized embeddings, and $\text{Pool}(\cdot)$ denotes using the embedding of `[CLS]` as sequence-level representation by following prior works. Given $\boldsymbol{v}$ of both head and tail events, we apply an interactive concatenation (Bowman et al., 2015; Reimers and Gurevych, 2019) between them to model their inferential relationship, i.e.,

$$\boldsymbol{r} := \text{Inter-Concat}(\boldsymbol{h}, \boldsymbol{t}) = [\boldsymbol{h}; \boldsymbol{h} \times \boldsymbol{t}; \boldsymbol{h} - \boldsymbol{t}; \boldsymbol{t}],$$
$$\text{where } \boldsymbol{h} = \boldsymbol{v}^{(h)} \text{ and } \boldsymbol{t} = \boldsymbol{v}^{(t)}. \quad (5)$$

Here, $\boldsymbol{r} \in \mathbb{R}^{4d}$ represents inferential relation between head and tail events, $[\cdot; \cdot]$ denotes vector concatenation, and "×" denotes element-wise product.

Lastly, the relation representation, $\boldsymbol{r}$, is learned by passing it into a neural classifier to predict the oracle relation in the original triple. In order to enable this knowledge model to represent a null or non-associated relation between events, we define an extra relation category, named dummy relation $r^{(dmy)}$. This classification is written as

$$\boldsymbol{p}^{(rc)} = \text{softmax}(\text{MLP}(\boldsymbol{r}; \theta^{(rc)})) \in \mathbb{R}^{|\mathcal{R}'|}, \quad (6)$$
$$\mathcal{R}' = \mathcal{R} \cup \{r^{(dmy)}\},$$

where $\boldsymbol{p}^{(rc)}$ is the probability distribution over $\mathcal{R}'$, and $\mathcal{R}'$ denotes a union of the well-defined relation set $\mathcal{R}$ with a dummy relation category $r^{(dmy)}$. The training data corresponding to $r^{(dmy)}$ is derived from negative sampling in the eventuality KG.

Training. We use a cross-entropy loss to optimize this discriminative knowledge model, $\{\theta^{(km)}, \theta^{(rc)}\}$, towards such a dummy-aware relation classification, which is denoted as

$$\mathcal{L}^{(rc)} = -\sum\nolimits_{(e^{(h)},r,e^{(t)})} \log \boldsymbol{p}^{(rc)}_{[y=r]}. \tag{7}$$

Inference. The trained knowledge model can be used in three ways summarized as (1) producing event representation by

$$\begin{aligned} \boldsymbol{v} &:= \text{Event-Enc}(e; \theta^{(km)}) \quad (8) \\ &= \text{Pool}(\text{Transformer}(\tilde{\boldsymbol{w}}^{e}; \theta^{(km)})), \end{aligned}$$

(2) generating relation representation by

$$\boldsymbol{r} := \text{Relation-Model}(e^{(h)}, e^{(t)}; \theta^{(km)}), \tag{9}$$

and (3) deriving a confidence score for whether there is an associated relation between two events:

$$\begin{aligned} p^{(ca)} &:= \text{Confid}(e^{(h)}, e^{(t)}; \theta^{(km)}, \theta^{(rc)}) \quad (10) \\ &= \sum\nolimits_{r \in \mathcal{R}' \backslash \{r^{(dmy)}\}} \boldsymbol{p}^{(rc)}_{[y=r]}, \end{aligned}$$

Remark. This discriminative knowledge model learns inferential relations between events in latent space, facilitating event-centric reasoning tasks. But it has its drawbacks like incompetence to auto-construction, in contrast to the generative knowledge models. Thereby, we argue generative and discriminative knowledge models are complementary to each other with different downstream uses.

3.2 Script-Adaptive Knowledge Model

In multi-choice narrative cloze (MCNC), a script reasoning model is asked to capture the relation between an event chain and a subsequent event candidate, however beyond the ability of the proposed knowledge model. To handle the MCNC task, we propose a neural adapter for the event encoder in Eq.(8), making it competent in modeling an event chain. Our goal is that, given a subsequent candidate, we extract the most relevant "event" from an event chain to represent the whole chain. As such, the result is still compatible with high-layer components in our knowledge model.

To this end, we present a chain-dependent attention module which is based on bidirectional chain contexts $\boldsymbol{e} = [e_1, \dots, e_n]$ queried by a potential subsequent event $e^{(c)}$. In particular, we first generate event representation for each event by our trained event encoder, i.e.,

$$\begin{aligned} \boldsymbol{v}^{(c)} &= \text{Event-Enc}(e^{(c)}; \theta^{(km)}), \quad (11) \\ \boldsymbol{V} &= [\boldsymbol{v}_1, \dots, \boldsymbol{v}_n] \in \mathbb{R}^{d \times n}, \quad (12) \\ &\text{where, } \boldsymbol{v}_i = \text{Event-Enc}(e_i; \theta^{(km)}), \end{aligned}$$

Then the embedded event chain, $\boldsymbol{V}$, position-wisely concatenated with the query event representation $\boldsymbol{v}^{(c)}$, is passed into a bidirectional long short-term memory (Bi-LSTM) to model rich event-contextual information of the event chain, i.e.,

$$\begin{aligned} \bar{\boldsymbol{U}} &= \text{Bi-LSTM}(\bar{\boldsymbol{V}}; \theta^{(bl)}) \in \mathbb{R}^{4d \times n}, \quad (13) \\ \boldsymbol{U} &= \text{MLP}(\bar{\boldsymbol{U}}; \theta^{(rd)}) \in \mathbb{R}^{d \times n}, \quad (14) \\ &\text{where, } \bar{\boldsymbol{V}} = [\bar{\boldsymbol{v}}_1, \dots, \bar{\boldsymbol{v}}_n], \bar{\boldsymbol{v}}_i = [\boldsymbol{v}_i; \boldsymbol{v}^{(c)}] \in \mathbb{R}^{2d}. \end{aligned}$$

The resulting $\boldsymbol{U} \in \mathbb{R}^{d \times n}$ is chain-dependent representations of the chain events, $\text{MLP}(\cdot; \theta^{(rd)})$ is responsible for reducing dimensionality.

Lastly, a self-attention pooling module (Liu et al., 2016; Lin et al., 2017) is applied to $\boldsymbol{U}$ to get a vector representation of the event chain, i.e.,

$$\begin{aligned} \boldsymbol{c} &= \boldsymbol{U}\boldsymbol{\alpha}, \quad (15) \\ &\text{where } \boldsymbol{\alpha} = \text{softmax}(\text{MLP}(\boldsymbol{U}; \theta^{(sa)})). \end{aligned}$$

Here, $\boldsymbol{\alpha} \in \mathbb{R}^{n}$ denotes the probability distribution of attention mechanism, which is then applied to chain-dependent event representations $\boldsymbol{U} \in \mathbb{R}^{d \times n}$ by matrix multiplication. As a result, $\boldsymbol{c}$ denotes a chain-dependent event representation extracted from the whole event chain. Intuitively, it can be viewed as the most relevant event from the event chain $\boldsymbol{e}$ to the candidate event $e^{(c)}$ as it is derived from an attention module queried by $e^{(c)}$. Hence, the derived $\boldsymbol{c}$ is still compatible with the top layers (e.g., interactive concatenation and neural classifier) in the discriminative knowledge model. Note that, the parameters of this neural script adapter, $\theta^{(adp)} = \{\theta^{(bl)}, \theta^{(rd)}, \theta^{(sa)}\}$, will be learned towards the MCNC objective jointly with other neural components in our script reasoning model, which is detailed in the next section (§3.3).

In summary, we can define a chain-dependent event encoding module to the above procedures to embed an event chain, i.e.,

$$\boldsymbol{c} = \text{Event-Enc}^{(\text{adp})}(\boldsymbol{e}, e; \theta^{(km)}, \theta^{(adp)}), \tag{16}$$

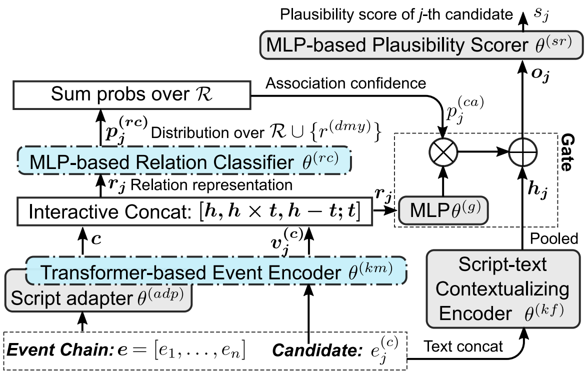
Plausibility score of j-th candidate s_j
MLP-based Plausibility Scorer $\theta^{(sr)}$
$\boldsymbol{o_j}$
Sum probs over $\mathcal{R}$
Association confidence
$p_j^{(ca)}$
$\boldsymbol{p}_j^{(rc)}$ Distribution over $\mathcal{R} \cup \{r^{(dmy)}\}$
MLP-based Relation Classifier $\theta^{(rc)}$
Gate
$\boldsymbol{r}_j$ Relation representation
$\boldsymbol{r}_j$
$\boldsymbol{h}_j$
Interactive Concat: $[\boldsymbol{h}, \boldsymbol{h} \times \boldsymbol{t}, \boldsymbol{h} - \boldsymbol{t}; \boldsymbol{t}]$
MLP$\theta^{(g)}$
$\boldsymbol{c}$
$\boldsymbol{v}_j^{(c)}$
Pooled
Transformer-based Event Encoder $\theta^{(km)}$
Script adapter $\theta^{(adp)}$
Script-text Contextualizing Encoder $\theta^{(kf)}$
Event Chain: $\boldsymbol{e} = [e_1, \ldots, e_n]$
Candidate: $e_j^{(c)}$
Text concat

Figure 3: Our script reasoning model for MCNC. Freezing the trained modules (dash-dot blue rounded rectangles) in our knowledge model, we optimize the other learnable modules (solid gray rounded rectangles) in our script reasoning model towards the MCNC task. Please refer to Figure 2 for an illustration of the script adapter.

where $\boldsymbol{e} = [e_1, \ldots]$ is an event chain and e is a query event. The chain-dependent event representation, $\boldsymbol{c}$, can be used as an argument to $\text{Inter-Concat}(\cdot, \cdot)$ to model script-level relationship with another event chain or a single event. Thus, the other two inference models in Eq.(9) and Eq.(10) are also adapted to $\text{Relation-Model}^{(\text{adp})}(\cdot)$ and $\text{Confid}^{(\text{adp})}(\cdot)$ respectively.

3.3 Script Reasoning Model

Built upon the discriminative knowledge model and its script adapter, we lastly present our script learning model for multi-choice narrative cloze task. To be specific, given an event chain $\boldsymbol{e} = [e_1, \ldots, e_n]$ and each event $e_j^{(c)}$ from the subsequent candidates $\mathcal{E}^{(c)}$, we first pass them into the script-adaptive knowledge model to generate a chain-dependent event representation $\boldsymbol{c}_j$ as defined Eq.(16):

$$\boldsymbol{c}_j = \text{Event-Enc}^{(\text{adp})}(\boldsymbol{e}, e_j^{(c)}; \theta^{(km)}, \theta^{(adp)}),$$

where $\forall e_j^{(c)} \in \mathcal{E}^{(c)}$. Based on $\boldsymbol{c}_j$, we can also derive the relation representation $\boldsymbol{r}_j$ between the event chain and each candidate, as well as the confidence score p_j of the association.

$$\boldsymbol{r}_j = \text{Relation-Model}^{(\text{adp})}(\boldsymbol{e}, e_j^{(c)}; \theta^{(km)}, \theta^{(adp)}),$$
$$p_j^{(ca)} = \text{Confid}^{(\text{adp})}(\boldsymbol{e}, e_j^{(c)}; \theta^{(km)}, \theta^{(rc)}, \theta^{(adp)}).$$

Besides the above rich-relation features from the knowledge model, we also leverage expressively powerful contextualized representations from another pre-trained bidirectional Transformer to fully exploit implicit reasoning knowledge in event texts. Formally, we present a *script-text contextualizing encoder* that applies the Transformer encoder to a concatenation of the event chain and each subsequent candidates, with special tokens separated:

$$\tilde{\boldsymbol{w}}_j = (\texttt{[CLS]}, \boldsymbol{w}^{e_1}, \ldots, \texttt{[SEP]}, \boldsymbol{w}^{e_j^{(c)}}, \texttt{[SEP]}),$$
$$\boldsymbol{h}_j = \text{Pool}(\text{Transformer}(\tilde{\boldsymbol{w}}_j; \theta^{(kf)})). \qquad (17)$$

To integrate the knowledge from the both models, we present a *knowledge gating module* for element-wise addition weighted by the association confidence:

$$\boldsymbol{o}_j = \boldsymbol{h}_j + p_j^{(ca)} \cdot \text{MLP}(\boldsymbol{r}_j; \theta^{(g)}), \qquad (18)$$

where, $\boldsymbol{o}_j \in \mathbb{R}^d$ is the final vector to represent the relation between the chain and a candidate from two perspectives, and $\text{MLP}(\cdot; \theta^{(g)})$ is responsible for reducing dimensionality from $4d$ to d. Such a gating module leads to a flexible knowledge integration, which is prone to avoiding redundant, non-associated relation features.

Finally, an MLP-based scoring module is defined to calculate a plausibility score given the final relation representation, followed by a softmax to derive predicted distribution:

$$s_j = \text{MLP}(\boldsymbol{o}_j; \theta^{(sr)}), \forall j = 1, \ldots, m, \qquad (19)$$
$$\boldsymbol{p}^{(sr)} = \text{softmax}([s_1; \ldots; s_m]) \in \mathbb{R}^m, \qquad (20)$$

where $m = |\mathcal{E}^{(c)}|$, and $\boldsymbol{p}^{(sr)}$ is the predicted distribution over candidate events $\mathcal{E}^{(c)}$ in MCNC.

Training. With fixed knowledge model $\{\theta^{(km)}, \theta^{(rc)}\}$, we train both the adapter $\theta^{(adp)}$ and the reasoning model $\theta^{(src)} = \{\theta^{(kf)}, \theta^{(g)}, \theta^{(sr)}\}$ towards the objective of MCNC, by a cross-entropy loss, i.e.,

$$\mathcal{L}^{(sr)} = -\frac{1}{|\mathcal{D}|} \sum_{\mathcal{D}} \log \boldsymbol{p}^{(sr)}_{[y=e^{(c)*}]}, \qquad (21)$$

where $e^{(c)*}$ denotes the gold subsequent event.

Inference. We can obtain the most plausible subsequent event from a trained MCNC model by

$$\hat{e}^{(c)} = \arg\max_{e_j^{(c)}} [s_1; \ldots; s_m]. \qquad (22)$$

4 Experiments

This section begins with a detailed description of our experimental setups on multi-choice narrative cloze (MCNC) task for script reasoning. Then, we conduct quantitative evaluations on the MCNC task, followed by extensive qualitative evaluations, including ablation study, model analysis, case study and error analysis.

Method	ACC (%)
w/o external knowledge	
Random	20.00
PMI (Chambers and Jurafsky, 2008)	30.52
Bigram (Jans et al., 2012)	29.67
Word2vec (Le and Mikolov, 2014)	42.23
Event-Comp (Granroth-Wilding and Clark, 2016)	49.57
PairLSTM (Wang et al., 2017)	50.83
SGNN (Li et al., 2018)	52.45
RoBERTa$_{\text{base}}$ (Lv et al., 2020b)	56.23
w/ external knowledge	
SGNN + Int&Senti (Ding et al., 2019)	56.03
RoBERTa$_{\text{base}}$ + Knwl. (Lv et al., 2020b)	58.66
Ours + RoBERTa$_{\text{base}}$	**59.99**
Ours + RoBERTa$_{\text{large}}$	**63.62**

Table 1: Comparison of our approach with previous script reasoning models on MCNC task. Our two models achieve 59.96% and 63.95% on dev set, respectively.

Datasets and Knowledge Graph. Following prior works (Lv et al., 2020b) for script reasoning, we evaluate our proposed approach on the dataset published by Li et al. (2018), which is widely used for the MCNC task. We follow the official data split[1] with 140,331/10,000/10,000 samples in training/dev/test sets. We use ASER (Zhang et al., 2020) as an external knowledge graph and learn a knowledge model from it. ASER is a large-scale eventuality knowledge graph extracted from unstructured textual data. It contains 15 event relations, 194M unique events, and 64M event-centric triples.

Evaluation Metrics. We adopt the official evaluation metric (Li et al., 2018), accuracy (ACC), to measure the performance of the reasoning models.

Implementation Details. The proposed approach for script reasoning contains two training processes, one for the discriminative knowledge model pre-trained on the eventuality KG and the other for the script reasoning model for MCNC task. (1) For the knowledge model, we adopt the BERT$_{\text{base}}$ model and optimize the cross-entropy loss with Adam optimizer. The learning rate is set to 1×10^{-5}. The hidden size of Bi-LSTM is set to 256. The maximum training epoch and batch size are set to 100 and 128. The maximum sequence length and dropout are set to 18 and 0.1. The weight decay and gradient clipping are set to 0.01 and 1.0. (2) For the script reasoning model, We experiment with two pre-trained language models, i.e. RoBERTa$_{\text{base}}$ and RoBERTa$_{\text{large}}$. Both the embedding size and hidden size are set to 768 in RoBERTa$_{\text{base}}$ and 1024 in RoBERTa$_{\text{large}}$. We use Adam optimizer (Kingma and Ba, 2015) to optimize the cross-entropy loss. The learning rate is set to 1×10^{-5}. The maximum training epoch and batch size are set to 3 and 32. The maximum sequence length and dropout are set to 64 and 0.1. The weight decay and gradient clipping are set to 0.01 and 1.0. We choose the model with the best result on the development set and report the results on the testing set are based on this model. The knowledge model contains 110M parameters, and our reasoning models contain 127M and 359M parameters for the base and large initializations, respectively. Our experiments are conducted on 4 NVIDIA P40 GPUs, and the training time is around 5 hours with RoBERTa$_{\text{base}}$ and 9 hours with RoBERTa$_{\text{large}}$.

Method	ACC (%)
Ours + RoBERTa$_{\text{large}}$	**63.62**
w/o chain-dependent attention	62.62
w/o knowledge gating module	62.85
w/o script-adapter	62.24
w/o external knowledge	61.53

Table 2: Ablation study of our approach. "*w/o chain-dependent attention*" denotes replacing chain-dependent attention module in our script adapter with mean-pooling, "*w/o knowledge gating module*" denotes removing confidence score $p^{(ca)}$ of the gating module in Eq.(18), ''*w/o script-adapter*" denotes ablating both chain-dependent attention and knowledge gating module, and "*w/o external knowledge*" denotes removing our script-adaptive knowledge model, equivalent to the RoBERTa$_{large}$ baseline.

4.1 Main Evaluation

The experimental results of our approach and previous script reasoning works on the Multi-Choice Narrative Cloze (MCNC) task are shown in Table 1. From the table, we can make two observations. First, using external knowledge, especially external event graph knowledge, increases the accuracy of models. For example, the knowledge infusion approach proposed by Lv et al. (2020b) outperforms the RoBERTa model without any knowledge. Second, our approach is superior to the *retrieval and integration* approach, RoBERTa + Knwl, and achieves new state-of-the-art accuracy (i.e. 63.62% using the RoBERTa$_{\text{large}}$ text encoder) on this task. This demonstrates the effectiveness of the proposed script-adaptive knowledge model.

4.2 Ablation Study

We conduct an ablation study to investigate the effect of each component of our approach and the results are reported in Table 2. We first investigate the impact of the chain-dependent attention

[1] https://github.com/eecrazy/ConstructingNEEG_IJCAI_2018

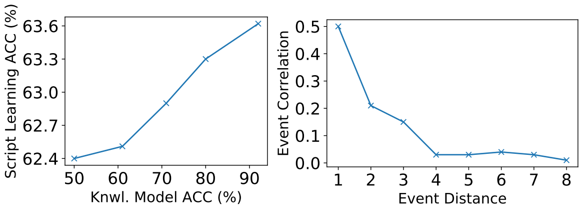

Figure 4: Impacts of knowledge model accuracy (left) and event distance over event association (right).

in Eq.(15) by replacing it with mean pooling over all events in the chain, and find that the accuracy of script reasoning is decreased about 1%. Next, we testify our approach without the knowledge gating module, which decreases the accuracy by 0.8%. And the gap becomes 1.4% if both the chain-dependent attention and the confidence score is ablated. Finally, we compare our approach with the baseline without any external knowledge included, and the accuracy of script reasoning drops by 2.1%, demonstrating the effectiveness of leveraging external event knowledge by the discriminative knowledge model.

4.3 Model Analysis

Impact of Knowledge Model. Our approach leverages the external event KG by directly modeling relations between event pairs. Intuitively, the accuracy of the learned model plays a critical role in script reasoning. Thus, we investigate its impact by assessing the performance of script reasoning with the knowledge models of various accuracy. The accuracy of knowledge model is evaluated on its dev set from the KG. As shown in Figure 4 (left), we can observe that the accuracy of script reasoning increases with that of the knowledge model, verifying the assumption that integrating external event knowledge via a knowledge model improves the performance of a script reasoning model.

Impact of Event Distance over Event Association. The script reasoning task requires to predict a subsequent event given an event chain. We investigate how the distance between two events impacts their correlation by analyzing the attention score of various timesteps of precedent events in a script. In particular, for all precedent events which are i-th step before subsequent events, we aggregate their attention scores. The results are plot in Figure 4 (right). The x-axis represents the distance between two events, and y-axis represents the estimated correlation by our model. From the figure we can see that an event is most highly correlated to its precedent neighbor. The association drops quickly as the distance increases, and it becomes very small when two events are three steps away.

Precedent Events
...
E7: state granted accountant
E8: believed accountant
Subsequent Event Candidates
A. asked accountant (*correct answer*)
B. ends accountant
C. depict accountant
D. penalize accountant firm
E. need accountant salary
Event pairs in KG with similar events
<Zakia will believe it, Zakia drops he guard>
<He is an accountant, He is very intelligent>
. . .
Event pairs in KG with Similar Relation Pattern
<I need your medical expertise, I need you help on something>
<You be the expert, I need answer>

Table 3: An example of script reasoning. The task is to choose a subsequent event from 5 candidates for the given precedent events. The associated event pair in the example is marked in blue with underline.

4.4 Case Study

As demonstrated in Table 3, we present an example in the test set to compare the *retrieval and integration* approach and ours. Here the script describes an event chain about *accountant*, which states that *accountants are believed*.

To infer the next event, the *retrieval and integration* paradigm will try to retrieve events with similar lexicon or semantics, e.g. "*zakia will believe it*", "*he is an accountant*", etc. However, KG triples containing these events do not capture the relation that *if a person is believed, people will consult him/her*. Therefore, this approach fails to leverage the KG to make a correct prediction.

In contrast, the KG contains event pairs like "*(I need your medical expertise, I need your help on something)*", "*(you are the expert, I need answer)*", whose relation patterns support the relation between "*accountants are believed*" and "*people ask accountants*", although there is little overlap between KG events and scrip events. The knowledge model learned from the KG event pairs captures such relation patterns and provides strong support for reasoning in this example, which demonstrates the effectiveness of our approach compared with the *retrieval and integration* approach.

4.5 Error Analysis

Lastly, to analyze the limitations of our model, we investigate the mis-classified examples on the MCNC test set, and summarize two main problems:

First, some scripts consist of precedent events which might lead to conflict results. For example, a event chain, ["*He disappointed supporters*", "*He fulfilled promise*"], is likely to be associated with two opposite results. The former might be associate with "*He lost campaign*", while the latter might result in "*They backed up his campaign*". Such case might confuse the reasoning model.

Second, long-distance dependency between events are difficult to capture. For example, in a tourism script which describes "*Emily went to the beach*" followed by a long description about the parking problem she met, although "*Emily went surfing*" is a rational subsequent event, the distance between the two events is too long so it is difficult for a reasoning model to capture such relations.

5 Related Work

A script (Schank and Abelson, 2013) refers to a kind of structured representation for prototypical sequences of events. Chambers and Jurafsky (2008) formulate a script learning (narrative learning) task and propose statistical models to capture event co-occurrence for subsequent event prediction. Afterwards, the approaches for script reasoning can be categorized into two genres. i.e., *event pair modeling* (Jans et al., 2012; Pichotta and Mooney, 2014; Granroth-Wilding and Clark, 2016) and *event chain modeling* (Pichotta and Mooney, 2016; Wang et al., 2017; Lv et al., 2019). But, they still lag far behind humans as the well-labeled training set is usually of small size. In addition, script reasoning is more challenging than traditional NLP tasks and requires models to reason over unobserved events.

With recent developments of large-scale eventuality knowledge graphs (KG) (e.g., ASER (Zhang et al., 2020) and ATOMIC (Sap et al., 2019)), an effective remedy is to adopt "*retrieval and integration*" schema and integrate the inferential facts retrieved from the KG for script reasoning (Lv et al., 2020b). This paradigm is proven effective in both entity-centric and concept-centric tasks, such as relation extraction (Zhang et al., 2019), named entity recognition (Liu et al., 2020) and commonsense reasoning (Lin et al., 2019; Lv et al., 2020a), etc. However, this paradigm is not that compatible with event-centric script reasoning since script reasoning focuses more on the inferential relation between consecutive events in a script rather than the triple facts with exact event matching. What is worse, these eventuality KGs consisting of free-form event usually encounter low knowledge coverage or incompleteness problem (Zhang et al., 2020; Bosselut et al., 2019b), leading to problematic grounding from an event to the nodes in the KG.

To circumvent the coverage problem, Bosselut et al. (2019b) and Hwang et al. (2020) propose to learn a *generative knowledge model* on existing triples from an eventuality KG, where the triples can be regarded as a seed of knowledge. It on-demand generates subsequent events with a prompt of the observed event and an inferential relation, thus avoiding event grounding and satisfying coverage necessity for a broad spectrum of NLP tasks (Shwartz et al., 2020; Majumder et al., 2020; Paul and Frank, 2020; Ding et al., 2019; Ma et al., 2019). However, such generative knowledge models are not perfectly compatible when capturing inferential relations between events because they focus more on inferring tail events rather than the relations.

In contrast, our method avoids operating merely on the triples that have lexical or semantic overlap with the targeted script, while directly learn the inferential relation patterns on the whole KG. The learned knowledge model can simply capture the relation between events in a script in latent space, benefiting various event-centric reasoning tasks.

6 Conclusion

In this work, we explore a novel paradigm to integrate an external eventuality knowledge graph into a script reasoning model for multi-choice narrative cloze task. We first identify a major problem affecting the integration for script reasoning. That is, previous works merely retrieve the graph triples that have semantic overlap with the events in a script, but neglect that the triples with similar inferential relation patterns can contribute a lot. We hence propose a knowledge model that learns the patterns on the graph and then provides supportive rich-relation evidence for events in a script. We also present a script adapter to make the knowledge model compatible with script-level reasoning. Built upon these, we finally present a reasoning model and evaluate it on the targeted task. Experimental results demonstrate that, the proposed model delivers new state-of-the-art performance, followed by further analyses to provide comprehensive insights.

References

Antoine Bosselut, Ronan Le Bras, and Yejin Choi. 2019a. Dynamic neuro-symbolic knowledge graph construction for zero-shot commonsense question answering. *arXiv preprint arXiv:1911.03876*.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Çelikyilmaz, and Yejin Choi. 2019b. COMET: commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 4762–4779. Association for Computational Linguistics.

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 632–642. The Association for Computational Linguistics.

Nathanael Chambers and Daniel Jurafsky. 2008. Unsupervised learning of narrative event chains. In *ACL 2008, Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, June 15-20, 2008, Columbus, Ohio, USA*, pages 789–797. The Association for Computer Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Xiao Ding, Kuo Liao, Ting Liu, Zhongyang Li, and Junwen Duan. 2019. Event representation learning enhanced with external commonsense knowledge. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 4893–4902. Association for Computational Linguistics.

Mark Granroth-Wilding and Stephen Clark. 2016. What happens next? event prediction using a compositional neural network model. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 2727–2733. AAAI Press.

Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. 2020. COMET-ATOMIC 2020: On symbolic and neural commonsense knowledge graphs. *CoRR*, abs/2010.05953.

Bram Jans, Steven Bethard, Ivan Vulic, and Marie-Francine Moens. 2012. Skip n-grams and ranking functions for predicting script events. In *EACL 2012, 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, April 23-27, 2012*, pages 336–344. The Association for Computer Linguistics.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Quoc V. Le and Tomás Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, volume 32 of *JMLR Workshop and Conference Proceedings*, pages 1188–1196. JMLR.org.

Zhongyang Li, Xiao Ding, and Ting Liu. 2018. Constructing narrative event evolutionary graph for script event prediction. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 4201–4207. ijcai.org.

Bill Yuchen Lin, Xinyue Chen, Jamin Chen, and Xiang Ren. 2019. Kagnet: Knowledge-aware graph networks for commonsense reasoning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 2829–2839. Association for Computational Linguistics.

Zhouhan Lin, Minwei Feng, Cícero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. 2017. A structured self-attentive sentence embedding. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. 2020. K-BERT: enabling language representation with knowledge graph. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 2901–2908. AAAI Press.

Yang Liu, Chengjie Sun, Lei Lin, and Xiaolong Wang. 2016. Learning natural language inference using bidirectional LSTM model and inner-attention. *CoRR*, abs/1605.09090.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Shangwen Lv, Daya Guo, Jingjing Xu, Duyu Tang, Nan Duan, Ming Gong, Linjun Shou, Daxin Jiang, Guihong Cao, and Songlin Hu. 2020a. Graph-based reasoning over heterogeneous external knowledge for commonsense question answering. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 8449–8456. AAAI Press.

Shangwen Lv, Wanhui Qian, Longtao Huang, Jizhong Han, and Songlin Hu. 2019. Sam-net: Integrating event-level and chain-level attentions to predict what happens next. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 6802–6809. AAAI Press.

Shangwen Lv, Fuqing Zhu, and Songlin Hu. 2020b. Integrating external event knowledge for script learning. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 306–315. International Committee on Computational Linguistics.

Kaixin Ma, Jonathan Francis, Quanyang Lu, Eric Nyberg, and Alessandro Oltramari. 2019. Towards generalizable neuro-symbolic systems for commonsense question answering. *CoRR*, abs/1910.14087.

Bodhisattwa Prasad Majumder, Harsh Jhamtani, Taylor Berg-Kirkpatrick, and Julian J. McAuley. 2020. Like hiking? you probably enjoy nature: Persona-grounded dialog with commonsense expansions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 9194–9206. Association for Computational Linguistics.

Debjit Paul and Anette Frank. 2020. Social commonsense reasoning with multi-head knowledge attention. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings, EMNLP 2020, Online Event, 16-20 November 2020*, pages 2969–2980. Association for Computational Linguistics.

Karl Pichotta and Raymond J. Mooney. 2014. Statistical script learning with multi-argument events. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2014, April 26-30, 2014, Gothenburg, Sweden*, pages 220–229. The Association for Computer Linguistics.

Karl Pichotta and Raymond J. Mooney. 2016. Learning statistical scripts with LSTM recurrent neural networks. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 2800–2806. AAAI Press.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 3980–3990. Association for Computational Linguistics.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, and Yejin Choi. 2019. ATOMIC: an atlas of machine commonsense for if-then reasoning. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 3027–3035. AAAI Press.

Roger C Schank and Robert P Abelson. 2013. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press.

Vered Shwartz, Peter West, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi. 2020. Unsupervised commonsense question answering with self-talk. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 4615–4629. Association for Computational Linguistics.

Reid Swanson and Andrew S. Gordon. 2008. Say anything: A massively collaborative open domain story writing companion. In *Interactive Storytelling, First Joint International Conference on Interactive Digital Storytelling, ICIDS 2008, Erfurt, Germany, November 26-29, 2008, Proceedings*, volume 5334 of *Lecture Notes in Computer Science*, pages 32–40. Springer.

Zhongqing Wang, Yue Zhang, and Ching-Yun Chang. 2017. Integrating order information and event relation for script event prediction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 57–67. Association for Computational Linguistics.

Hongming Zhang, Xin Liu, Haojie Pan, Yangqiu Song, and Cane Wing-Ki Leung. 2020. ASER: A large-scale eventuality knowledge graph. In *WWW '20:*

The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020, pages 201–211. ACM / IW3C2.

Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. ERNIE: enhanced language representation with informative entities. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 1441–1451. Association for Computational Linguistics.