Hybrid Generative and Commonsense Knowledge for Script Event Prediction

Anonymous ACL submission

Abstract

Script event prediction aims to predict subsequent events given contextual events, which requires inferring correlations between contexts and candidate events. Current research focuses on improving script event prediction using external knowledge and pre-trained language models, but faces the problems of sparse event-level correlation knowledge and separation of word-level correlation knowledge. In this paper, we propose a novel model CoGen-Predictor based on hybrid generative and commonsense knowledge that combines explicit event-level and implicit word-level correlation knowledge for prediction. CoGen-Predictor constructs event-level correlations through a commonsense knowledge base and updates the event representations using graph neural networks, then learns word-level contextual event correlations through a generative approach. Experimental results on the multi-choice narrative cloze (MCNC) task demonstrate the effectiveness of the model.

1 Introduction

Scripts (Schank & Abelson, 2013) refer to a type of structured knowledge that consists of a structured sequence of events. Figure 1 shows a restaurant dining script that involves a sequence of events that occur when a customer enters the restaurant. The script event prediction task (Granroth-Wilding & Clark, 2016) aims to select the correct subsequent events from the candidate events. Studying this task can gain event knowledge from the event chain and benefit many downstream tasks such as story generation (Chaturvedi et al., 2017), dialogue generation (Danescu-Niculescu-Mizil & Lee, 2011), and is also useful for studying reasoning in Large Language Models (LLMs).

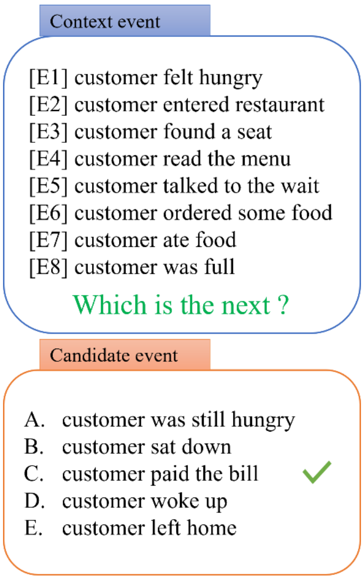

Figure 1: A simplified example of script event prediction.

Understanding events and inferring correlations between events is essential for acquiring knowledge of events and reasoning about subsequent events. In Figure 1, it is necessary to understand the contextual events and use the common knowledge that "need to pay for the meal" to infer that "customer paid the bill". Therefore, knowledge of explicit correlations between events and implicit correlations between words is essential: explicit correlations represent correlations between different contextual events, and implicit correlations represent correlations between words in the chain of events.

Recent work incorporates external knowledge to enhance models to understand event relevance, and while these approaches have yielded promising results, however, some challenges remain. First, current work suffers from sparsity when incorporating external knowledge. Specifically, relationship construction mainly relies on find-and-match, and such approaches often suffer from the problem that event knowledge cannot be fully adapted to the knowledge in the event knowledge base when injecting external knowledge bases, and the commonly used ASER (Zhang et al., 2020) knowledge base is actually a knowledge base of probabilistic event relationships, but in real reasoning it requires common sense knowledge that includes both discourse and correlation relations including a commonsense knowledge base. Secondly, current work modelling relevance uses a single approach of pre-trained language models, ignoring the complementarity of representation learning and generative paradigms in modelling relevance.

In order to better modelling the correlation between events and to introduce correlation knowledge into script event prediction tasks, this paper proposes CoGen-Predictor (Commonsense and Generative Predictor), a novel prediction model based on hybrid generative and commonsense knowledge. At the event-level explicit correlation knowledge level, CoGen-Predictor consists of a relation builder component and a representation learning component, which introduces explicit correlations by constructing event relations and updates event representations. At the word-level implicit correlation knowledge level, a pre-trained generative language model is used for fine-tuning and modelling implicit correlation between words. It simultaneously utilises event-level representation learning and word-level generative paradigm's correlation knowledge for subsequent event judgments. The main contributions of this paper are as follows:

- We introduce a commonsense knowledge base and train an event-relationship builder to model explicit correlations between events to solve the problem of sparse event-level relationships.

- We propose CoGen-Predictor, a novel model based on hybrid generative and commonsense knowledge, for scripted event prediction using event-level explicit correlation knowledge from builder modelling and textual implicit correlation knowledge from generative modelling.

- The experimental results of multi-choice narrative cloze (MCNC) show that the method in this paper can effectively utilise the two types of correlation knowledge and reduce the dependence on the original text to obtain state-of-the-art results.

2 Related Work

Script event prediction was first proposed by Chambers and Jurafsky (2008), who defined an event as a verb and its dependency and proposed the basic structure of a narrative event chain and a narrative completion task. Subsequently, Granroth-Wilding and Clark (2016) extended the event definition to the verb and three theses〈subject; objecti; indirect object〉, and proposed the currently widely used multi-choice narrative cloze (MCNC) task: selecting subsequent events from candidate events based on contextual narrative event chains.

Earlier work (Granroth-Wilding & Clark, 2016) obtained event representations via Word2Vec, and then aggregated pointwise mutual information (PMI) between contextual script events and candidate events to infer the probability that a candidate event is a subsequent event of the script. Earlier studies ignored the narrative order of scripted events, and subsequent works (Lv et al., 2019; Wang et al., 2017) introduced LSTM to integrate temporal information between events.

Many works have been conducted to model the relationship between events, including the use of graph structure (Zhongyang et al., 2018) and external knowledge (Ding et al., 2019) to improve event representations. Recent studies include Gao et al. (2022) who proposed a method for learning event representations with both weakly supervised contrast learning and clustering. Du et al. (2022) introduced a pre-trained model, BERT (Devlin et al., 2018), and replaced the model middle layer with a graph neural network to embed the eventgraph information. Lee, Pacheco, and Goldwasser (2020) mined discourse relations from the original text by template matching, Bai et al. (2021) enriched event representations using the original text, Lv et al. (2020); Zhou et al. (2021) introduces event knowledge graph ASER (Zhang et al., 2020) to augment the pre-trained model RoBERTa (Liu et al., 2019); and Wang et al. (2023) uses R-GCN to learn the correlation information between events.

Recent researchers (Zhou, Geng, et al., 2022; Zhou, Shen, et al., 2022) perform event-centred pre-training on an external corpus and employ an event-level masking strategy (Zhu et al., 2023) to fine-tune the generative model to generate predicted subsequent events.

The above methods can be divided into two main categories: inter-event relationship modelling and candidate event generation. Inter-event relationship modelling faces the problem of sparse relationship construction, and candidate event generation faces the problem of using pre-trained models in a single way. The CoGen-Predictor proposed in this paper overcomes these challenges by effectively introducing a wide range of discourse and event relations through an event relation builder, avoiding the sparsity problem that arises in relation modelling, combining event-level explicit correlation knowledge and word-level implicit correlation knowledge from generative modelling, and performs well in the task of scripted event prediction.

3 Preparations

Problem Statement. As shown in Figure 1, script event prediction is defined as predicting the most likely subsequent event for a given script. Formally, given a script $X = \{x_1, x_2, \ldots, x_n\}$ and a candidate event $Y = \{y_1, y_2, \ldots y_m\}$, where x_i and y_j represent the events, this task aims at selecting the correct subsequent event y_t from Y. Each event $e = (e_v, e_s, e_o, e_i)$ consists of a predicate e_v and three arguments (subject e_s, object e_o and indirect object e_i). The model needs to compute the relevance score $P(y_j|x)$ for each candidate event $y_j (j \in 1, \ldots, m)$ given a script X, and then take the most likely event as the subsequent event.

Commonsense Knowledge Graph. CKGs are large-scale knowledge bases that store knowledge in a graph structure, focusing on the association of things or objects. in this paper, we introduce the DISCOS (Fang et al., 2021) commonsense knowledge base, which converts discourse knowledge about events from the large-scale discourse knowledge graph ASER (Zhang et al., 2020) to ATOMIC (Sap et al., 2019) to if-then commonsense knowledge defined in ATOMIC (Sap et al., 2019), which provides 3.4 million inference commonsense knowledge and event knowledge, providing more effective support for constructing event relations in SEP.

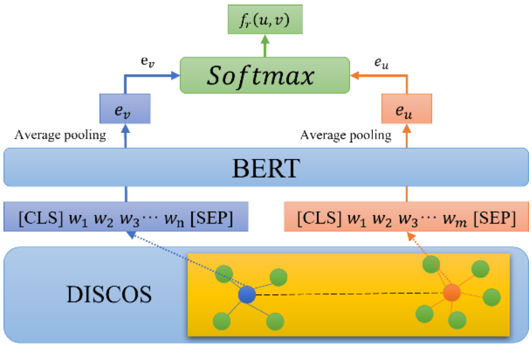

Figure 2: The training process of the event relationship builder.

4 The CoGen-Predictor Model

In this section, the CoGen-Predictor model of this paper is described. As shown in Fig. 2, it has four main components: (1) event relationship builder, (2) event-level explicit correlation representation modelling, (3) word-level implicit correlation generative modelling and (4) fusion prediction scoring.

4.1 Event Relationship Builder

For the nodes in the DISCOS commonsense knowledge base G_r^c, all its nodes are encoded using the pre-trained language representation model BERT. For nodes $v = [w_1, w_2, \ldots, w_n]$ with n-length tokens, a [CLS] token is added at the beginning of each sentence as w_0, and a [SEP] token is added at its end as w_{n+1}. Represent the output of the input BERT as $[e_{w_0}, e_{w_1}, \cdots, e_{w_{n+1}}]$, $e_{w_i} \in \mathbb{R}^d$, where d is the embedding dimension of the BERT. For the tuple $(u, v) \in G_r^c$ for making a relational judgement, the semantic representation $[e_u, e_v]$ is obtained using the BERT layer, and the final correlation score is obtained using the Softmax layer:

$$f_r(u, v) = Softmax([e_u, e_v]W'^T + b), \qquad (1)$$

Here $W' \in \mathbb{R}^{2 \times d}$, $b \in \mathbb{R}^2$, and after obtaining the predicted value of $f_r(u, v)$, the correlation is judged to be constructed or not according to the score, where $p = argmax(f_r(u, v))$.

Each node in DISCOS is an event, and in order to better match the commonsense knowledge in DISCOS, contextual events in the Script Event Prediction (SEP) task is first formalised into a matching form.

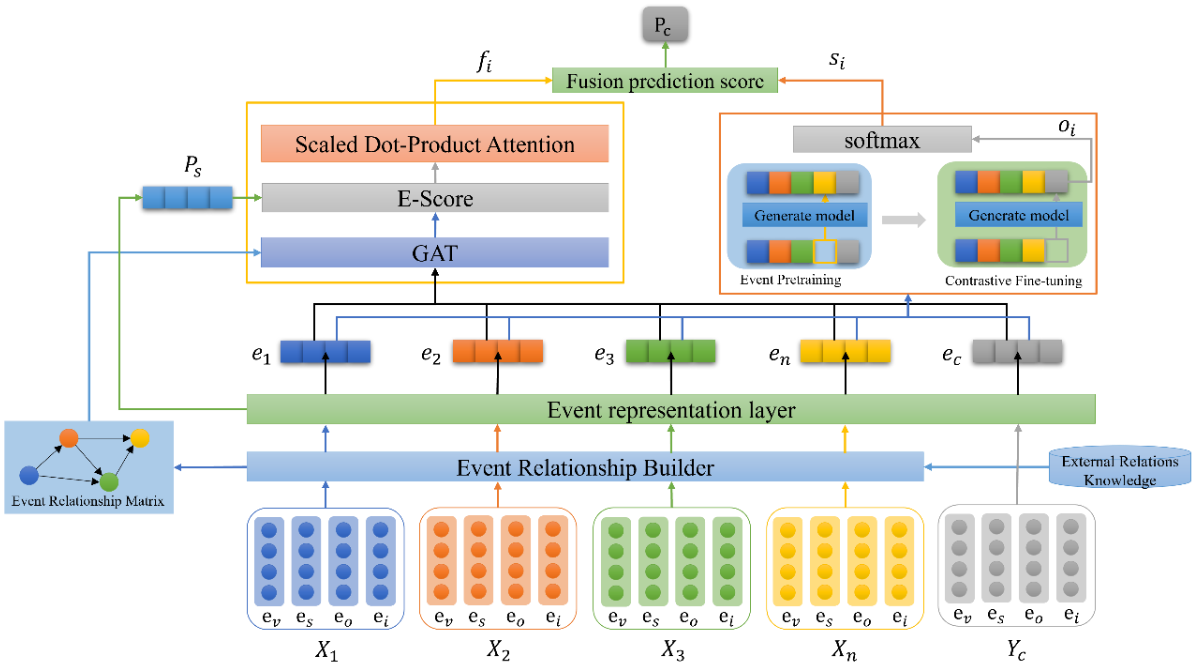

Figure 3: Structure of the CoGen-Predictor model.

Given the head and tail events e^h and e^t, each event consists of four elements $e = (e_v, e_s, e_o, e_i)$, so adjust the order of discourse to represent the event as " e_s e_v e_o e_i " to make it match the commonsense knowledge format. The natural language text w^e for each event e is linked with special tags:

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

The textual representations $\widetilde{w}^{e^h}$ and $\widetilde{w}^{e^t}$ of the head and tail events e^h and e^t are then imported into the Event Relationship Builder to obtain the result of the relationship construction R.

4.2 Event-level representation modelling

Based on the relationship R, CoGen-Predictor uses BART for semantic representation and incorporates GAT to fuse event-level explicit correlation knowledge for modelling event and event chain information.

4.2.1 Event Representation

Using the pre-trained model BART (Lewis et al., 2020) as the underlying semantic representation model, for the input context events $X = \{x_1, \cdots, x_n\}$ and candidate events $Y = \{y_1, \cdots, y_m\}$, each consisting of $e = (e_v, e_s, e_o, e_i), e \in X, Y$, denote the events as "$e_s$ e_v e_o e_i" to adjust the semantic order to input BART, the context event sequence S is denoted as $< s > x_1 < SEP > x_2 < SEP > \cdots < SEP > x_n </s>$, and each candidate event is denoted as $< s > y_m </s>$, with the candidate events being placed into the list C. The sequence of contextual events S and the list of candidate events C,are input into the BART model for text encoding:

$$P_s, P_c = BART(S, C), \quad (3)$$

$P_s, P_c \in \mathbb{R}^d$ and d is the embedding dimension of BART.

4.2.2 Related Information Update

Based on the relation R , CoGen-Predictor introduces the graph attention neural network (GAT) for node updating of context event nodes. The first step is to partition the events in the context event sequence into independent event representations:

$$\{x_1, x_2, \cdots, x_n\} = segmentation(P_s), \quad (4)$$

Linear transformation is performed on each input node feature to obtain the mapped feature and the attention coefficient e_{ij} is calculated for each pair of neighbouring nodes i and j:

$$h'_i = W \cdot x_i, \quad (5)$$

$$e_{ij} = LeakyReLU(\alpha^T [W \cdot x_i || W \cdot x_j]) \quad (6)$$

where W is the weight matrix, $||$ denotes the connection operation, α is the weight vector, and $LeakyReLU$ is the activation function.

The steps for event node update using GAT (Velickovic et al., 2017) are as follows:

Attention coefficients of all neighbours of node i are normalised using $softmax$ function:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k\in \mathrm{N}(i)} \exp(e_{ik})}, \tag{7}$$

where $\mathrm{N}(i)$ denotes the set of i's neighbours. Weighted summation of neighbouring node features using normalised attention coefficients to update node features:

$$h_i'' = \sigma\left(\sum_{j\in N(i)} \alpha_{ij} W \cdot x_j\right), \tag{8}$$

where σ is a nonlinear activation function (e.g., ReLU). Here the contextual event node information is updated by GAT:

$$\{x''_1, x''_2,\cdots, x''_n\} = GAT(\{x_1, x_2,\cdots, x_n\}), \tag{9}$$

4.2.3 Candidate Event Scoring

In order to apply both event information and event chain information, we include the obtained event chain text representation P_s when constructing the context node representation, so the final context representation is:

$$H = \{x''_1, x''_2,\cdots, x''_n\}||P_s, \tag{10}$$

The widely used negative Euclidean distance is used here as a score calculation for candidate events:

$$s_i = -\left|\left|h_i - h_j\right|\right| ** 2, \tag{11}$$

Here h_i comes from H after updating the node and h_j comes from P_c which has been encoded by BART.

Different contextual events and event chains contribute differently to predicting the correct candidate event, we use scaled dot product attention (Vaswani et al., 2017) to aggregate the distance scores of different nodes:

$$f = \sum_{i=0}^{n-1} \alpha_i s_i\,, \tag{12}$$

After the above representation modelling the score of each candidate event relative to the event node and event chain is obtained.

4.3 Word-level generative modelling

Referring to the approach of Zhu et al. (Zhu et al., 2023), CoGen-Predictor employs a generative model to model the knowledge of implicit correlations between textual words, which is divided into two main phases: event pre-training and comparison fine-tuning.

4.3.1 Event pre-training

In the event-centred pre-training phase, the script and the correct candidate event y_t are concatenated into an event sequence $S = \{x_1, x_2,\cdots, x_n, y_t\}$. Mask K events in S, where $K \in (1,2,3)$, and the masked events are denoted as the event sequence E. Denote the contextual event sequence S as $< s > x_1 . x_2 . < MASK > . \cdots . < MASK > . \cdots . y_t </s>$, where the masking event is replaced with the token $< MASK >$, and the event sequence E is represented in a similar way. Using the generative model BART as the backbone, the conditional probability distribution $P(E|S)$ is formulated as follows:

$$P(E|S) = \frac{1}{N_E}\sum_{n=2}^{N_E} logP_{LM}(E^n|S, E^{1:n-1})\,, \tag{13}$$

where N_E is the token number of the event sequence E in natural language format, E^n is the nth token, and $E^{1:n-1}$ is the first to n-1th token of E.

In the event pre-training phase, the CoGen-Predictor generative paradigm part is trained to maximise $P(E|S)$.

4.3.2 Contrast fine-tuning

The modified sequence is represented as X_m by first adding a marker $[MASK]$ at the end of the script X. X_m and each event candidate X_m are then converted to a natural language format using the average of the log probabilities of the descriptive markers for each event as the score o_i for the event y_i.

$$o_i = \frac{1}{N_{y_i}}\sum_{n=2}^{N_{y_i}} logP_{LM}({y_i}^n|X_m, {y_i}^{1:n-1})\,, \tag{14}$$

where N_{y_i} is the length of event y_i. Then, the Softmax function is used to calculate the final score s_i for each candidate event y_i:

$$s_i = \frac{\exp(o_i)}{\sum_{k=1}^{M} \exp(o_k)}, \tag{15}$$

Finally, define the loss function as follows:

$$\mathcal{L}_{cot} = -log(s_t) + \frac{1}{M-1}\sum_{\substack{i=1 \\ i\neq t}}^{M} \left(\frac{s_i}{1-s_t}\right) log\left(\frac{s_i}{1-s_t}\right), \tag{16}$$

where t is the subscript of the right event candidate y_t and M is the number of event candidates.

4.4 Integration predictions

Both event-level explicit correlation knowledge and word-level implicit correlation knowledge are used for subsequent event judgements. To integrate the knowledge from the two components, CoGen-

Public Dataset	Number
Train set	140,331
Dev set	10,000
Test set	10,000

Table 1: Dataset.

Predictor uses a gating module that sums the elements weighted by association confidence. The previous explicit knowledge scoring result f_i and the implicit knowledge scoring result s_i are obtained, where $i \in (0, m-1)$:

$$F_i = s_i + p \cdot f_i, \tag{17}$$

For the obtained final score F_i, we selects the most likely event with the highest score as the predicted subsequent event y_p , where $p = argmax(F_i)$.

5 Experiments

In this section, CoGen-Predictor is compared to some baselines to validate its effectiveness. In addition, an ablation study is performed to understand the impact of key components of the model on performance. Finally, a case study is performed to demonstrate how the model in this paper predicts subsequent events.

5.1 Dataset

In the task of script event prediction, most of the existing work selects the public dataset published by Li, Ding, and Liu (2018). Therefore, this public dataset is also used in this paper. We follows the common practice of dataset segmentation used for training, validation and testing in Table 1. For the public dataset, each instance has five candidate events, of which only one choice is correct.

5.2 Experimental setup

The CoGen-Predictor model proposed in this paper includes both explicit correlation knowledge and implicit correlation knowledge construction. To compare with the baseline, experiments are conducted on $BART_{base}$ and $BART_{Large}$. The models were optimised by Adam (Kingma & Ba, 2015). The learning rate and weight decay are 1e-5 and 1e-6, respectively. the model in this paper uses an early stopping strategy to select the best epoch, and the patience is set to 5. For $BART_{base}$, the Batch Size is set to 32, and for $BART_{Large}$, the Batch Size is set to 24. all the experiments are carried out on the RTX 4090D. The GPU training time for event relationship builder module, event-level explicit association representation modelling and text-level implicit association generative modelling are about 8, 10 and 6.5. We select the model with the best results on the validation set and report the results on the test set, using accuracy as the evaluation metric.

Method	Acc.(%)
Random	20.00
Event-Comp	49.57
Pair-LSTM	50.83
SGNN	52.45
SAM-Net	55.60
GraphBERT	60.72
SGNN + Int&Senti	56.03
RoBERTa_{base} + Rep. Fusion	58.66
RoBERTa_{base} + Know.Model	59.99
BART_{base} + Contrastive Approcah	62.22
CoGen-Predictor (BART_{base})	**63.54**

Table 2: Base model comparison experiment.

Method	Acc.(%)
BART_large + Contrastive	63.40
EventBERT	63.50
RoBERTa_{large} + Know.Model	64.62
ClarET	64.61
MCPredictor-s	59.24
MCPredictor	**67.14**
CoGen-Predictor ($BART_{large}$)	**65.01**

Table 3: Large model comparison experiment.

5.3 Baselines

For the preliminaries, they can be divided into three categories. Event representation method: **1)Event-Comp** (Granroth-Wilding & Clark, 2016) uses training objectives such as Word2Vec to learn event embeddings and compute pairwise similarities between scripted events and candidate events. **2)Pair-LSTM** (Wang et al., 2017) uses LSTM to model the narrative order of script events. **3)SAM-Net** (Lv et al., 2019) uses LSTM and self-attention mechanisms to capture different event fragments. **4)MCPredictor** (Bai et al., 2021) obtains event representations from pre-trained Word2Vec and augments them with raw sentence representations obtained from pre-trained BERT, and uses multiple similar event chains to aggregate script-level information. **5)MCPredictor-s** is an ablation of MCPredictor that removes additional raw sentence information.

Structured information enhancement: **1)RoBERTa + Rep. Fusion** (Lv et al., 2020) integrates external knowledge from episodic knowledge graphs, ASER (Zhang et al., 2020), and predicts using RoBERTa. **2)RoBERTa + Know.Model** (Zhou et al., 2021) learns knowledge models from ASER to predict event relationships. **3)SGNN** (Zhongyang et al., 2018) constructs narrative event evolution graphs through verb co-occurrence frequencies for more effective event representation. **4)SGNN + Int & Senti** incorporates external intent and affective knowledge from ATOMIC (Sap et al., 2019) into event representations. **5)GraphBERT** (Du et al., 2022) constructs an event graph similar to SGNN and enhances BERT with the event graph.

Event-centred pre-training: **1)EventBERT** (Zhou, Geng, et al., 2022) pre-trains RoBERTa in BOOKCORPUS (Zhu et al., 2015) with three self-supervised comparative learning objectives: correlation-based event ranking, contradictory event labelling, and discourse relation ranking. **2)ClarET** (Zhou, Shen, et al., 2022) pre-trains BART on BOOKCORPUS with three additional self-supervised goals: overall event recovery, comparative event-related coding, and cue-based event localisation. **3)BART + Contrastive Approcah** (Zhu et al., 2023) uses both an event-centred pre-training phase and a task-specific contrastive fine-tuning phase for training.

5.4 Results and analyses

In this paper, the $BART_{base}$ and $BART_{Large}$ models are used as the backbone for training and testing in Tables 2 and 3, respectively, in order to ensure that the parameters of the comparison models remain relatively consistent, and the following observations are made from the results of the two models:

- CoGen-Predictor shows a 3.55% and 1.32% improvement over the best baseline RoBERTa + Know.Model for combining external knowledge and the best baseline BART + Contrastive Approcah for generative modelling of event relationships, respectively. CoGen-Predictor combines the GAT model's modelling of external commonsense correlational information for GAT model modelling and word-level correlational information for generative paradigm modelling, using external correlational knowledge in combination with knowledge of correlations within the event chain to enhance the ability to model correlational relationships between events.

Method	**Acc.(%)**
CoGen-Predictor	63.54
w/o GAT & Know	62.61
w/o Representing learning	62.20
w/o Generation method	59.36

Table 4: ablation experiment.

- The accuracy performance of CoGen-Predictor outperforms strong baselines after performing extensive event-centred pre-training, such as ClarET (Zhou, Shen, et al., 2022) and EventBERT (Zhou et al., 2021). Moreover, CoGen-Predictor is more advantageous in terms of training time and complexity compared to the previous two methods.

- CoGen-Predictor performs better in terms of accuracy relative to MCPredictor-s without using the original sentence text, and slightly inferior to MCPredictor. For MCPredictor extracting the original sentence of the event is crucial in the training process, and the accuracy of SCPredictor-s relative to MCPredictor with a 7.9% decrease in accuracy after removing the original text.CoGen-Predictor improves the model's effectiveness on the SEP task without using the original content through a more generalised DISCOS commonsense knowledge base with more generalisation capabilities.

5.5 Ablation experiment

Table 4 shows the results of the ablation experiments for the CoGen-Predictor model. Firstly, the effect of external knowledge base and GAT node updates is verified (row 1), and it is found that the performance decreases by 0.93% when the external correlation knowledge of GAT is missing. This is because the external commonsense correlation knowledge and GAT node updates provide explicit correlation knowledge that is not available in the event representation. Secondly, removing the event-level explicit association modelling component (line 2) resulted in a 1.34% decrease in performance because the lack of the event representation learning module made the

P	Acc.(%)
0.9	62.41
0.8	62.65
0.7	63.02
0.6	63.42
0.5	63.51
0.4	**63.54**
0.3	63.36
0.2	63.06
0.1	62.94

Table 5: Parametric studies.

event similarity understanding insufficient. Finally, the removal of the word-level implicit correlation generation modelling component (line 3) proves to be crucial for the SEP task as it comprehensively models inter-word implicit correlation.

5.6 Parametric studies

The aim of this section is to investigate how the parameter p of the fusion prediction layer affects the predictive performance of the model. The parameter p represents the fusion ratio between event-level explicit correlation information and word-level implicit correlation information. The prediction results under different parameters are shown in Table 5. From the results, it can be found that text-level implicit associative generative modelling contributes more in the final score compared to external commonsense correlation knowledge modelling. This may be due to the fact that word-level implicit correlations are more suitable for the SEP task compared to those generated based on the external commonsense knowledge base of correlations, but the external commonsense correlation knowledge injected in the model in this paper is also essential.

5.7 Case Studies

The case study demonstrates the properties of CoGen-Predictor and its predictive ability. The G-Score, I-Score, and F-Score in Table 6 represent the scores for generative models, external knowledge representation learning, and gated aggregation, respectively. In Case 1, the correct option "simpson attack deaths" does not have the highest score in the G-Score because the generative model can only handle word-level associations and lacks external knowledge support. In Case 2, the correct option "johnson share Monday gold" does not score well in the I-Score because the external knowledge alone does not allow for effective differentiation of textually relevant options. Combining event-level explicit and word-level implicit associations, F-Score correctly identifies subsequent events, demonstrating the effectiveness of CoGen-Predictor.

6 Conclusion

In this paper, we propose a novel hybrid generative and commonsense knowledge model, CoGen-Predictor, for script event prediction, which combines event-level explicit knowledge and word-level implicit knowledge and outperforms other state-of-the-art baseline models in the MCNC task. Future research will aim to incorporate external knowledge to better exploit the potential of generative models.

Events Case 1	**G-Score**	**I-Score**	**F-Score**	**Contextual event**
simpson come up rule	0.092	0.0267	0.1009	'friends know simpson', 'simpson become famous', 'simpson re', 'simpson kill friday', 'simpson write press', 'simpson use room', 'convict simpson', 'law apply simpson'
simpson invent defens	0.3479	0.1656	0.4142	
simpson circulate	0.3240	0.0987	0.3634	
simpson attack deaths	**0.1886**	**0.6448**	**0.4465**	
simpson pitch	0.0494	0.0642	0.0751	
Events Case 2	**G-Score**	**I-Score**	**F-Score**	**Contextual event**
accord johnson	'0.1194	0.0098	0.1233	'johnson finish end', 'johnson break record', 'johnson better record', 'johnson redefine event', 'johnson lament', 'johnson contract food', 'johnson advance', 'johnson claim gold'
lawyer bite johnson	0.0308	0.0943	0.0685	
johnson share monday gold	**0.4826**	**0.2672**	**0.5895**	
johnson be weak	0.3500	0.4908	0.5464	
johnson recruit	0.0171	0.1379	0.0722	

Table 6: Case Studies.

Limitations

Although our proposed method performs well on publicly available SEP datasets, it still suffers from several major limitations. Firstly, our method has requirements on the version format of the input data, which needs to have the verbs and their dependencies, and there may be a decrease in accuracy for problems with missing some parameters. Second, the dataset used in this paper is the standard dataset proposed in 2018, which may have poor portability due to the fact that only one dataset was used for the experiments. Third, our model uses BART as the backbone and may suffer from insufficient ability to generalise when dealing with specific linguistic contexts, which requires more cross-linguistic validation and adaptation tests. In addition, our experimental setup assumes the stability of data distribution, but changes in data distribution in real applications may affect the performance of the model. Finally, although our method performs well on the script event prediction task, its effectiveness in handling unstructured data needs to be further explored and improved.

References

Bai, Long, Saiping Guan, Jiafeng Guo, Zixuan Li, Xiaolong Jin, and Xueqi Cheng. 2021. "Integrating Deep Event-Level and Script-Level Information for Script Event Prediction." In 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, November 7, 2021 - November 11, 2021, 9869-78. Virtual, Punta Cana, Dominican republic: Association for Computational Linguistics (ACL).

Chaturvedi, Snigdha, Haoruo Peng, and Dan Roth. 2017. "Story comprehension for predicting what happens next." In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 1603-14.

Danescu-Niculescu-Mizil, Cristian, and Lillian Lee. 2011. "Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs." In 2nd Workshop on Cognitive Modeling and Computational Linguistics, CMCL 2011 at the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, ACL-HLT 2011, June 23, 2011, 76-87. Portland, OR, United states: Association for Computational Linguistics (ACL).

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. 'Bert: Pre-training of deep bidirectional transformers for language understanding', arXiv preprint arXiv:1810.04805.

Ding, Xiao, Kuo Liao, Ting Liu, Zhongyang Li, and Junwen Duan. 2019. "Event representation learning enhanced with external commonsense knowledge." In 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, November 3, 2019 - November 7, 2019, 4894-903. Hong Kong, China: Association for Computational Linguistics.

Du, Li, Xiao Ding, Yue Zhang, Kai Xiong, Ting Liu, and Bing Qin. 2022. "A Graph Enhanced BERT Model for Event Prediction." In 60th Annual Meeting of the Association for Computational Linguistics, ACL 2022, May 22, 2022 - May 27, 2022, 2628-38. Dublin, Ireland: Association for Computational Linguistics (ACL).

Fang, Tianqing, Hongming Zhang, Weiqi Wang, Yangqiu Song, and Bin He. 2021. "DISCOS: bridging the gap between discourse knowledge and commonsense knowledge." In Proceedings of the Web Conference 2021, 2648-59.

Gao, Jun, Wei Wang, Changlong Yu, Huan Zhao, Wilfred Ng, and Ruifeng Xu. 2022. "Improving Event Representation via Simultaneous Weakly Supervised Contrastive Learning and Clustering." In 60th Annual Meeting of the Association for Computational Linguistics, ACL 2022, May 22, 2022 - May 27, 2022, 3036-49. Dublin, Ireland: Association for Computational Linguistics (ACL).

Granroth-Wilding, Mark, and Stephen Clark. 2016. "What happens next? event prediction using a compositional neural network model." In Proceedings of the AAAI Conference on Artificial Intelligence.

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

Lee, I-Ta, Maria Leonor Pacheco, and Dan Goldwasser. 2020. "Weakly-supervised modeling of contextualized event embedding for discourse relations." In Findings of the Association for Computational Linguistics: EMNLP 2020, 4962-72.

Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2020. "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension." In 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, July 5, 2020 - July 10, 2020, 7871-80. Virtual, Online, United states: Association for Computational Linguistics (ACL).

Liu, Yinhan, 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." In.: arXiv.

Lv, Shangwen, 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 Proceedings of the AAAI conference on artificial intelligence, 6802-09.

Lv, Shangwen, Fuqing Zhu, and Songlin Hu. 2020. "Integrating external event knowledge for script learning." In Proceedings of the 28th International Conference on Computational Linguistics, 306-15.

Sap, Maarten, 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 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Annual Conference on Innovative Applications of Artificial Intelligence, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, January 27, 2019 - February 1, 2019, 3027-35. Honolulu, HI, United states: AAAI Press.

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

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. 'Attention is all you need', Advances in neural information processing systems, 30.

Velickovic, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. 'Graph attention networks', stat, 1050: 10-48550.

Wang, Zhongqing, Yue Zhang, and Ching-Yun Chang. 2017. "Integrating order information and event relation for script event prediction." In 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, September 9, 2017 - September 11, 2017, 57-67. Copenhagen, Denmark: Association for Computational Linguistics (ACL).

Wang, Zikang, Linjing Li, and Daniel Zeng. 2023. 'Integrating Relational Knowledge With Text Sequences for Script Event Prediction': 1-12.

Zhang, Hongming, Xin Liu, Haojie Pan, Yangqiu Song, and Cane Wing-Ki Leung. 2020. "ASER: A Large-scale Eventuality Knowledge Graph." In 29th International World Wide Web Conference, WWW 2020, April 20, 2020 - April 24, 2020, 201-11. Taipei, Taiwan: Association for Computing Machinery.

Zhongyang, Li, Ding Xiao, and Liu Ting. 2018. "Constructing narrative event evolutionary graph for script event prediction [C]." In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 4201-07.

Zhou, Yucheng, Xiubo Geng, Tao Shen, Guodong Long, and Daxin Jiang. 2022. "Eventbert: A pre-trained model for event correlation reasoning." In Proceedings of the ACM Web Conference 2022, 850-59.

Zhou, Yucheng, Xiubo Geng, Tao Shen, Jian Pei, Wenqiang Zhang, and Daxin Jiang. 2021. 'Modeling event-pair relations in external knowledge graphs for script reasoning', Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021.

Zhou, Yucheng, Tao Shen, Xiubo Geng, Guodong Long, and Daxin Jiang. 2022. "ClarET: Pre-training a Correlation-Aware Context-To-Event Transformer for Event-Centric Generation and Classification." In 60th Annual Meeting of the Association for Computational Linguistics, ACL 2022, May 22, 2022 - May 27, 2022, 2559-75. Dublin, Ireland: Association for Computational Linguistics (ACL).

Zhu, Fangqi, Jun Gao, Changlong Yu, Wei Wang, Chen Xu, Xin Mu, Min Yang, and Ruifeng Xu. 2023. "A Generative Approach for Script Event Prediction via Contrastive Fine-Tuning." In 37th AAAI Conference on Artificial Intelligence, AAAI 2023, February 7, 2023 - February 14, 2023, 14056-64. Washington, DC, United states: AAAI Press.

Zhu, Yukun, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. "Aligning books and movies: Towards story-like visual explanations by watching movies and reading books." In 15th IEEE International Conference on Computer Vision, ICCV 2015, December 11, 2015 - December 18, 2015, 19-27. Santiago, Chile: Institute of Electrical and Electronics Engineers Inc.