Appendix

Model	Posterior event Sequence	Accuracy
RoBERTa	-	73.2
ege-RoBERTa		
	$X' = \{O_1, H_i, O_2\}$	75.1
	$X' = \{O_1, \boldsymbol{I_1}, H_i, O_2\}$	77.1
	$X' = \{O_1, H_i, \boldsymbol{I_1}, O_2\}$	76.3
	$X' = \{O_1, \boldsymbol{I_1}, \boldsymbol{I_2}, H_i, O_2\}$	76.6
	$X' = \{O_1, H_i, \boldsymbol{I_1}, \boldsymbol{I_2}, O_2\}$	75.8
	$X' = \{O_1, \boldsymbol{I_1}, H_i, \boldsymbol{I_2}, O_2\}$	**77.9**

Table 1: Prediction accuracy of the ege-RoBERTa-base model pretrained with different forms of posterior event sequence.

1 Training Details

The ege-RoBERTa model is pretrained on the pseudo instance set for one epoch and finetuned on the αNLI dataset for three epochs. In both stage the learning rate are set to be 2e-5.

2 Influence of the Form of Posterior Event Sequence

In this paper, to equip ege-RoBERTa with event graph knowledge, without loss of generality, we arbitrary formalize the posterior event sequence as $X' = \{O_1, I_1, H_i, I_2, O_2\}$. Whereas our approach can also handle with other forms of posterior event sequence such as $X' = \{O_1, H_i, I_1, O_2\}$, which describes another possible relationship between the observed events, hypothesis event and intermediary event(s).

We enumerate possible forms of posterior event sequence in Table 1, and conduct experiments to investigate the specific relationship between forms of posterior event sequence and reasoning performance. All experiments are conducted on the dev set of αNLI using ege-RoBERTa-base. Results are shown in Table 1.

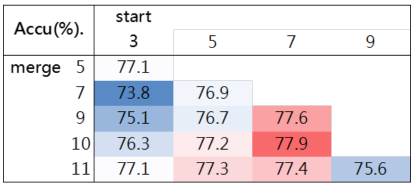
From Table 1 we can observe that:

(1) Compared with vanilla RoBERTa model, under all settings of posterior event sequence, ege-RoBERTa model show improvements in reasoning accuracy. This confirms that the event graph knowledge can be helpful for the abductive reasoning task. In addition, it also demonstrates the flexibility of our approach, as it can handle with various forms of posterior event sequence to enhance the performance of reasoning.

(2) Compared to setting $X' = \{O_1, H_i, O_2\}$, involving in at least one intermediary into the posterior event sequence can further improve the performance of reasoning. This indicates that in most case there exist intermediary events between the observed events and the hypothesis event. While ege-RoBERTa show most performance improvement when setting $X' = \{O_1, \boldsymbol{I_1}, H_i, \boldsymbol{I_2}, O_2\}$. This shows the reasonability of formalizing X' as $X' = \{O_1, H_i, I_1, O_2\}$.

3 Influence of the Start and Merge Layer

Since different transformer layers of RoBERTa tend to focus on different semantic and syntactic information (Clark et al., 2019; Coenen et al., 2019), as a result, which layer in RoBERTa is selected to aggregate event representations, and which layer is selected to merge the latent variable can affect the performance of the model. We study such effect through comparing the prediction accuracy of models with different (start layer, merge layer) combinations. Results are shown in Figure 1, from which we could observe that, employing the 7th transformer layer of RoBERTa as the start layer, and the 10th or 11th layer as the merge layer can achieve a higher prediction accuracy. Interestingly, Jawahar et al. (2019) find that syntactic features can be well captured in the middle layers of RoBERTa, especially in 7–10th layer. This indicates that middle layers of RoBERTa focus more on sentence level information, and implicitly support the reasonableness that choosing the 7th and 10th transformer



Accu(%).		start 3	5	7	9
merge	5	77.1			
	7	73.8	76.9		
	9	75.1	76.7	77.6	
	10	76.3	77.2	77.9	
	11	77.1	77.3	77.4	75.6

Figure 1: Prediction accuracy of models with different (start, merge) layer on development set of αNLI.

layer of RoBERTa as the start and merge layer.

References

Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D Manning. 2019. What does bert look at? an analysis of bert's attention. In *ACL Workshop*, pages 276–286.

Andy Coenen, Emily Reif, Ann Yuan, Been Kim, Adam Pearce, Fernanda Viégas, and Martin Wattenberg. 2019. Visualizing and measuring the geometry of bert. *arXiv preprint arXiv:1906.02715*.

Ganesh Jawahar, Benoît Sagot, and Djamé Seddah. 2019. What does bert learn about the structure of language? In *ACL 2019-57th Annual Meeting of the Association for Computational Linguistics*.