Improving Implicit Discourse Relation Recognition with Semantics Confrontation

Mingyang Cai, Zhen Yang, Ping Jian

Introduction

- Discourse relation recognition, aiming to identify the logical relation between two arguments, is a crucial task in discourse parsing.

- It encompasses two distinct paradigms: **Explicit** Discourse Relation Recognition (EDRR) and **Implicit** Discourse Relation Recognition (IDRR).

EDRR vs. **IDRR**

EDRR

Arg1: We're offering this plan now

Conn: [Because]

Arg2: We feel it's the right time

Relation Sense: Contingency

IDRR

Arg1: Living there for six years was really scary

Arg2: The ghosts of the past are everywhere

Relation Sense: Contingency

For EDRR, there are connectives (e.g., *because*, *so*) between the two discourse arguments, offering valuable prompts for the models to reason the semantic relations, while IDRR infers logical relation without the help of connectives.

Motivation

Type	arg_1	arg_2
Original	Living there for six years was really scary	(because) The ghosts of the past are everywhere
Same logic	Living there was scary	(because) The ghosts are everywhere
Diff logic	Living there for six years was really scary	(although) The ghosts of the past are gone

- The logical relation is totally different when replacing just one word (*everywhere* → *gone*)
- But the relation sense of the arguments still remains consistent when altering many other words.

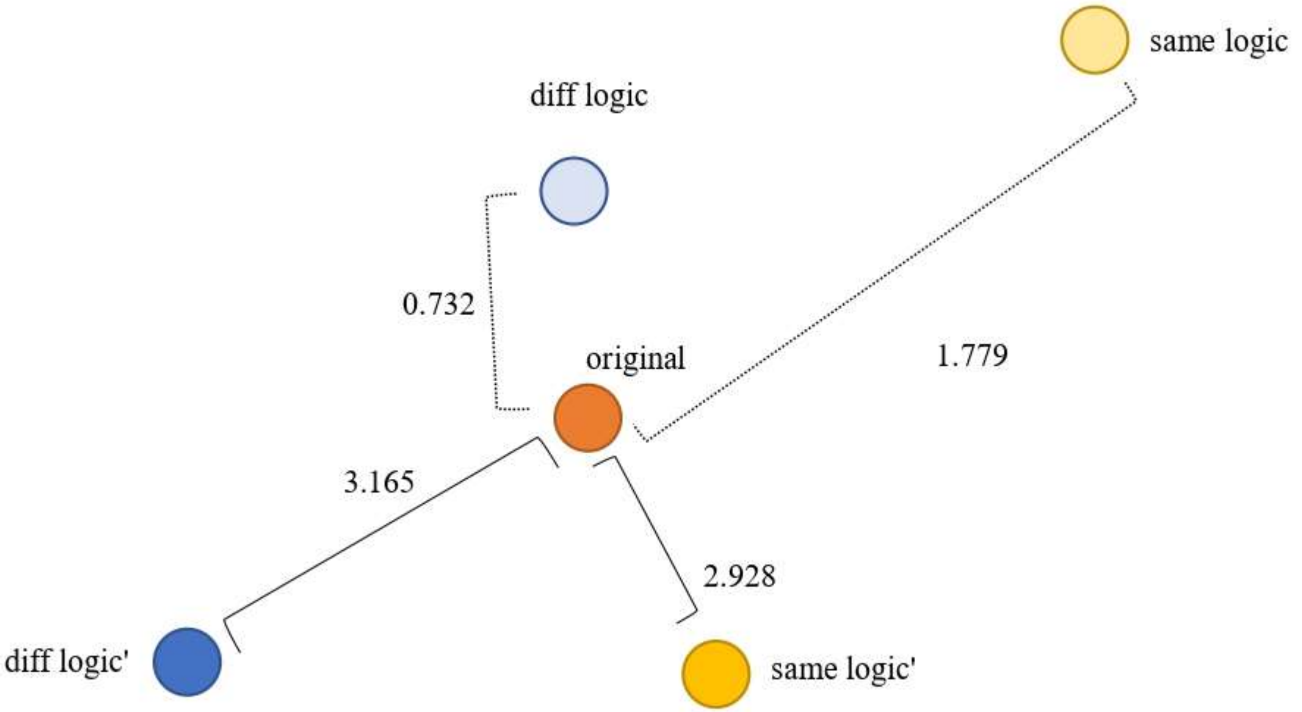
Type	arg_1	arg_2
Original	Living there for six years was really scary	(because) The ghosts of the past are everywhere
Same logic	Living there was scary	(because) The ghosts are everywhere
Diff logic	Living there for six years was really scary	(although) The ghosts of the past are gone

- We assume that any text unit consists of more than one aspect of semantics. As for IDRR, a robust discourse relation recognition model should be able to pronouncedly disentangle the logical semantics from logical-independent semantics (distinguished as "general semantics")
- Intuitively, the words contributing to logical semantics (denoted as logical words) are typically fewer in number compared to those contributing to general semantics (denoted as general words).

- Many pretrained language models (PLMs) employ mask language modeling (MLM) as pretraining task.
- Given that the distribution of logical words are more sparse than general ones, the masking process mainly involves general words.
- Result in a predominant focus on **learning the representation of general semantics, mixing with limited logical semantics**.

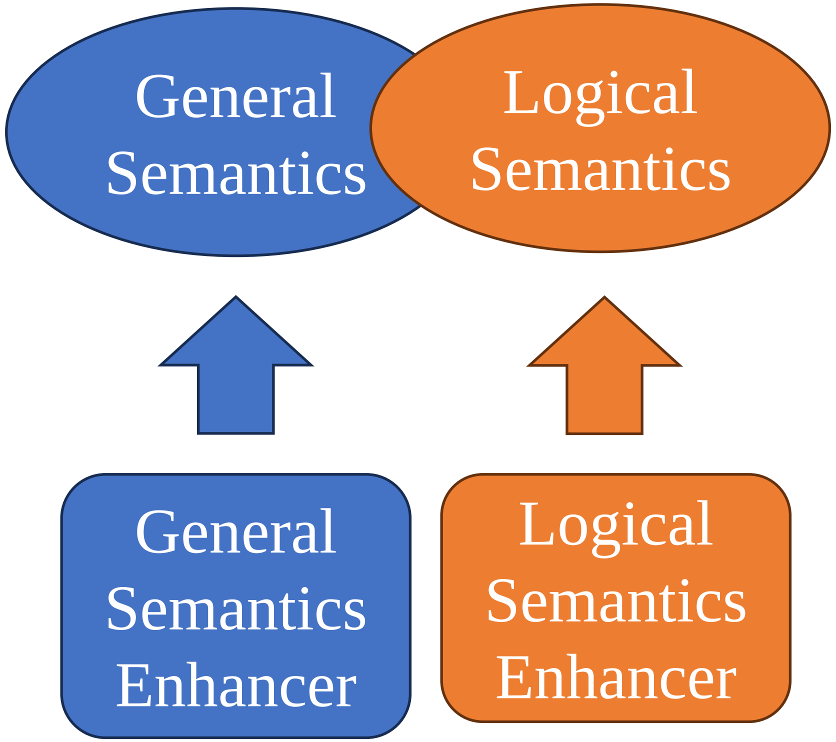
Verification

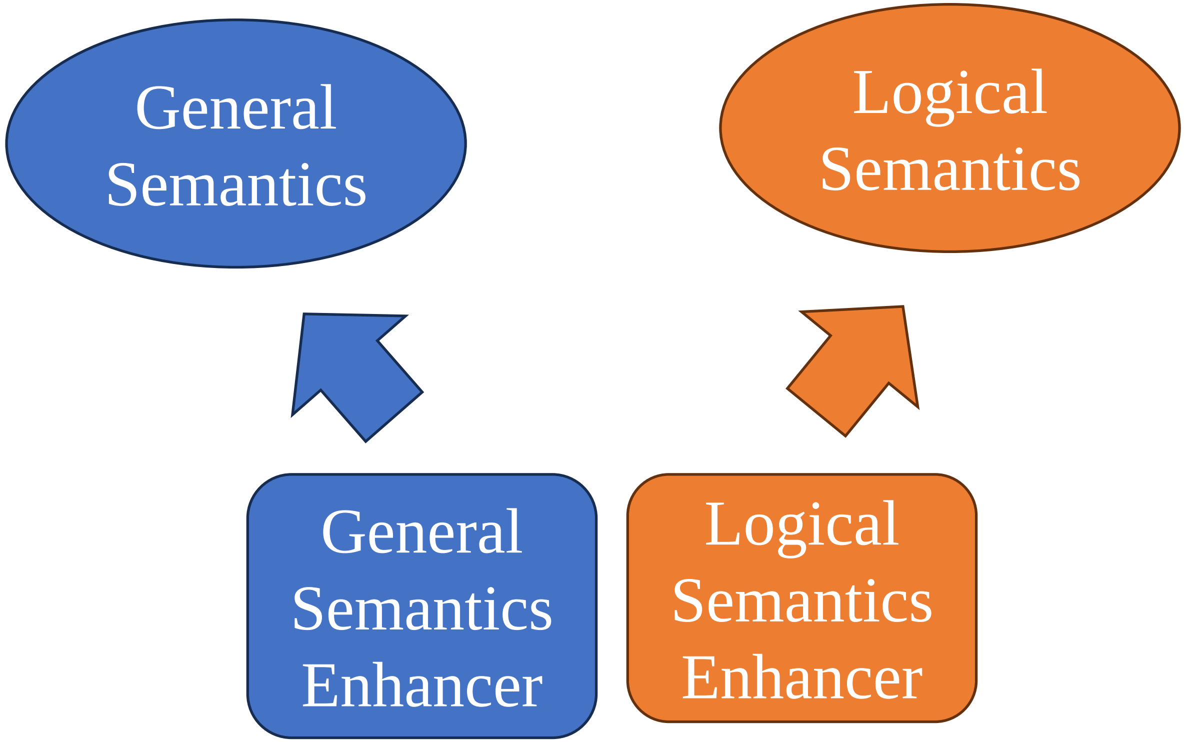
Type	arg_1	arg_2
Original	Living there for six years was really scary	(because) The ghosts of the past are everywhere
Same logic	Living there was scary	(because) The ghosts are everywhere
Diff logic	Living there for six years was really scary	(although) The ghosts of the past are gone



For embeddings obtained by PLMs,(*diff logic* & *same logic*) sentences exhibiting different logical relations are even closer to the original sentences than those sharing the same logical relations.

For embeddings obtained by our model, (*diff logic'* & *same logic'*) the sentences with same logic is closer than that with different logic.


General Semantics
Logical Semantics
General Semantics Enhancer
Logical Semantics Enhancer


General Semantics
Logical Semantics
General Semantics Enhancer
Logical Semantics Enhancer

Model Structure

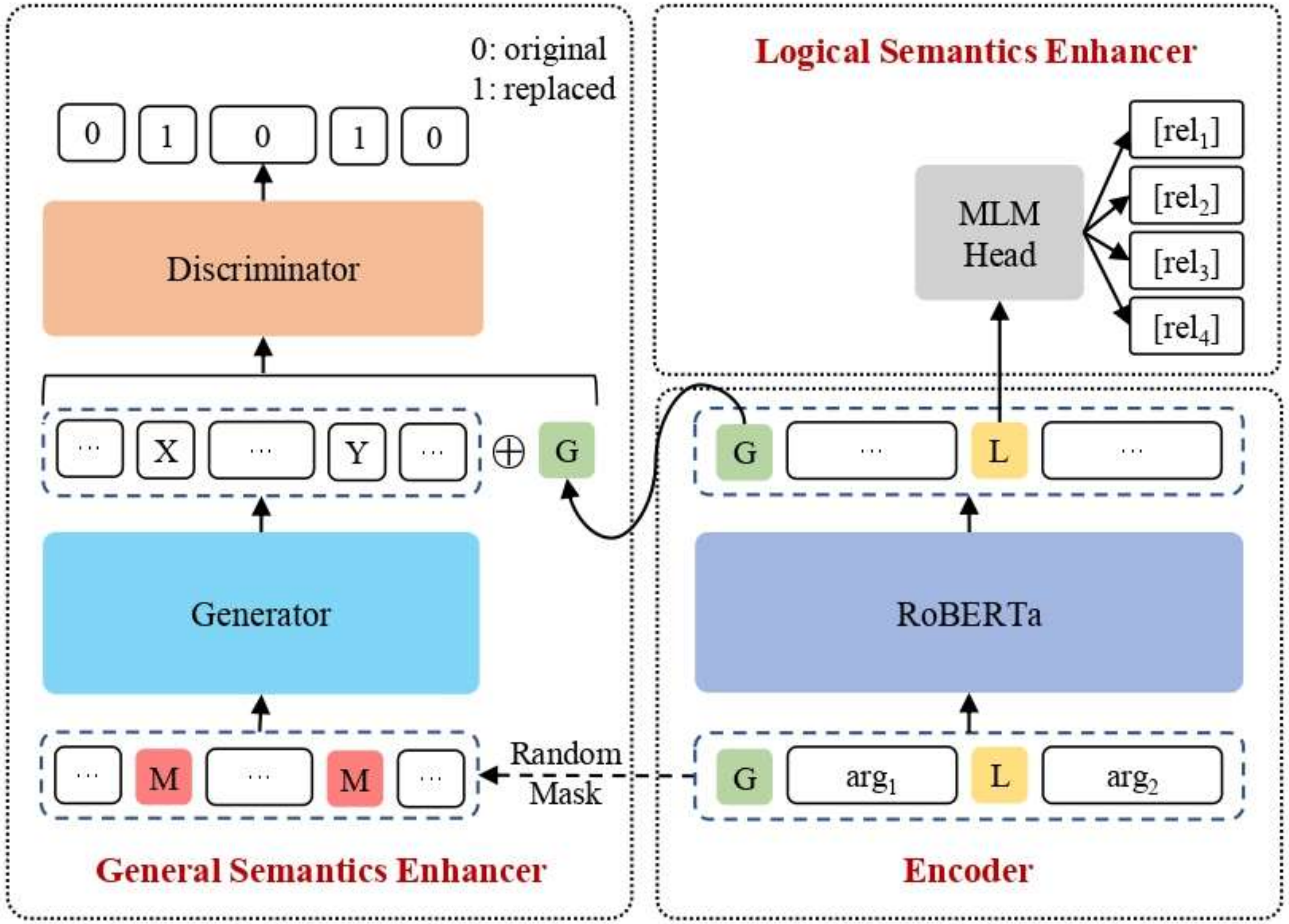
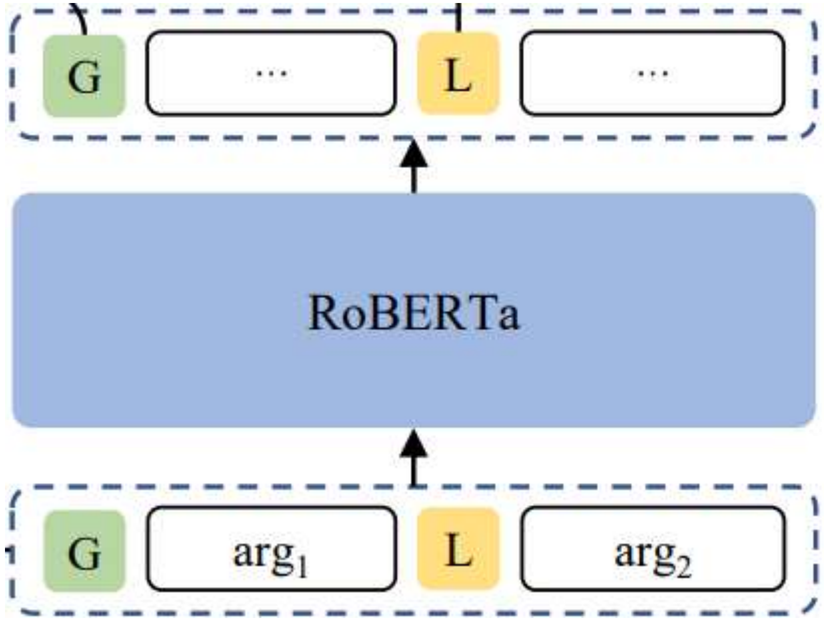


Figure 2: The model structure of our approach. The "G" represents "$[general]$" token, "L" represents "$[logical]$" token, "M" represents "$[MASK]$" token and $[rel_i]$ is manually designed token for each candidate logical relation, in order to be suitable for masked token predicting paradigm.

The model consists of three parts:

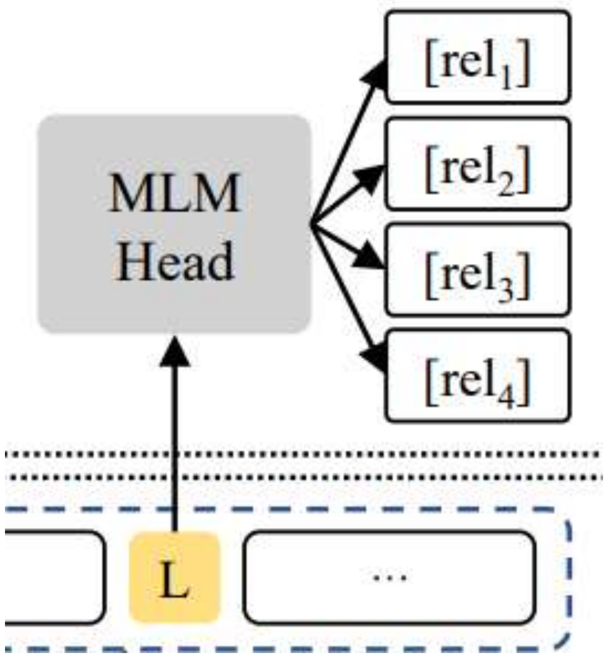
- Encoder
- Logical Semantics Enhancer
- General Semantics Enhancer

Encoder



- Use pretrained **RoBERTa** as encoder.
- Input: [*general*], Arg1, [*logical*], Arg2
- [*general*]: capture the global, long-term, and "dense" general semantics
- [*logical*]: align more closely with the natural language expression and capture logical semantics

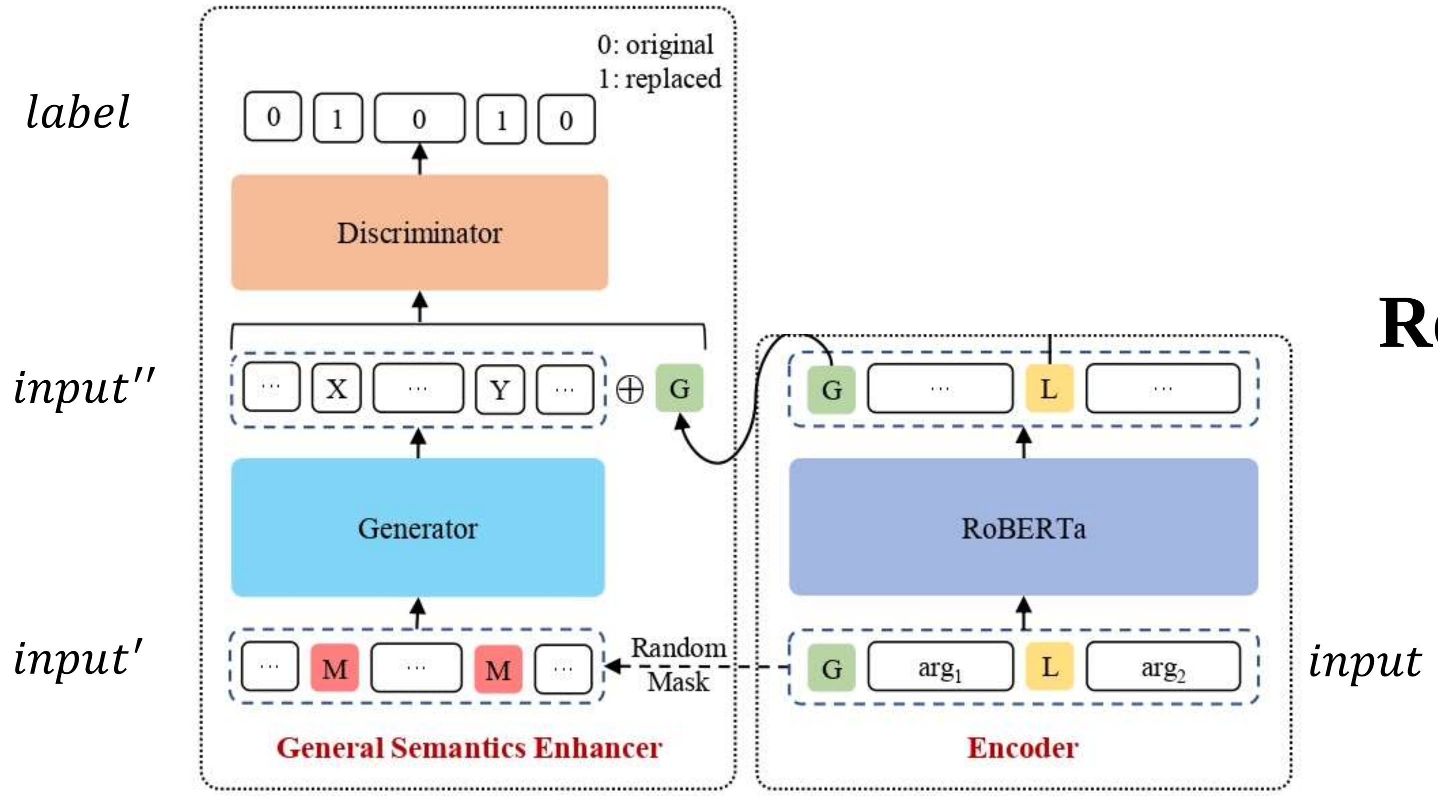
Logical Semantics Enhancer



Relation Token Prediction (RTP)

- Predict relation token $[rel_i]$ using MLM head
- Each relation token represents a specific relation sense

General Semantics Enhancer



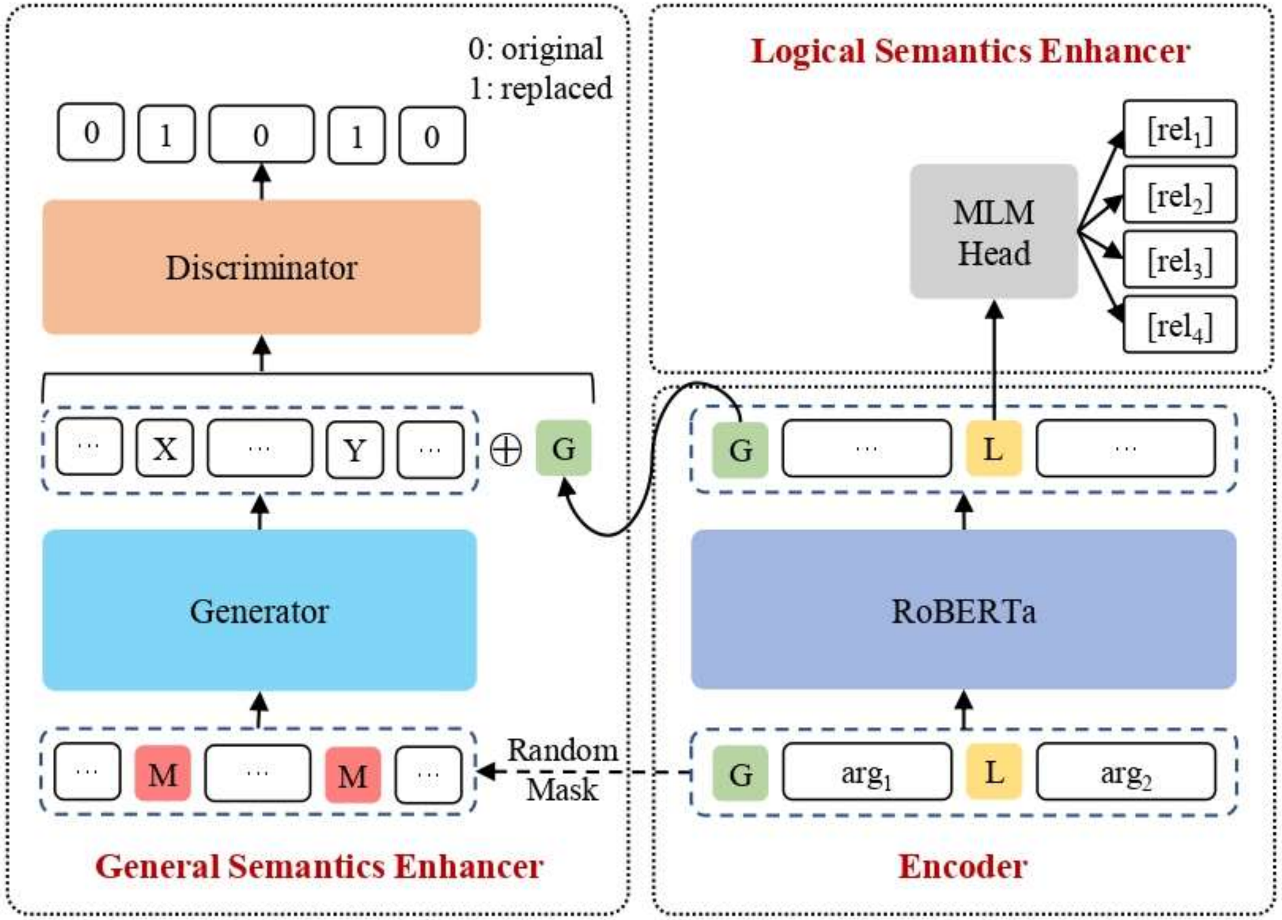
Type	Sentence			
input	The	ghosts	are	everywhere.
input'	The	$[MASK]$	$[MASK]$	everywhere
input''	The	cats	are	everywhere.
label	0	1	0	0

Table 3: The different content of sentences after passing through the general semantics enhancer.

Replaced Token Detection (RTD)

- *input* is randomly masked to get *input'*
- Generator produces *input''*
- Discriminator predicts whether *input''* is consistent with *input*

Training



- We train the general semantics enhancer after logical semantics enhancer during one epoch.
- In other words, two semantics enhancers are trained alternately.

Semantics Confrontation

- The two semantics enhancers inherently confront with each other.
- We leverage the inherent confrontation between the targets of RTP and RTD, thus training the two enhancers alternately to guide the representation of both tokens towards the logical and general semantics space, respectively.
- conduct implicit confrontation at the *representation level*

Main Results

Method	PDTB-4		PDTB-11	CoNLL-Test		CoNLL-Blind	
	F1	ACC	ACC	F1	ACC	F1	ACC
NNMA (Liu and Li, 2016)	46.29	57.17	-	-	-	-	-
Gshare (Lan et al., 2017)	47.80	57.39	-	-	39.40	-	40.12
Bi-LSTM-DU (Dai and Huang, 2018)	48.82	57.44	-	-	-	-	-
ELMo-C&E (Dai and Huang, 2019)	52.89	59.66	-	-	-	-	-
RWP-CNN (Varia et al., 2019)	50.20	59.13	-	-	39.39	-	39.36
KANN (Guo et al., 2020)	47.90	57.25	-	-	-	-	-
TransS (He et al., 2020)	51.24	59.94	-	-	-	-	-
BERT-HierMTN-CRF (Wu et al., 2020)	55.72	65.26	52.34	-	-	-	-
BERT-FT (Kishimoto et al., 2020)	58.48	65.26	54.32	-	-	-	-
BMGF-RoBERTa (Liu et al., 2020)	63.39	69.06	58.13	40.68	*57.26*	28.98	*55.19*
CG-T5 (Jiang et al., 2021)	57.18	65.54	53.13	-	-	-	-
CVAE-IDRR (Dou et al., 2021)	65.06	70.17	-	-	-	-	-
LDSGM (Wu et al., 2022)	63.73	71.18	60.33	-	-	-	-
PCP (Zhou et al., 2022b)	64.95	70.84	**60.54**	*33.27*	*55.48*	*26.00*	*50.99*
ChatGPT (Chan et al., 2023a)	36.11	44.18	24.54	-	-	-	-
Our Baseline Model	63.19	69.12	59.34	39.32	52.21	27.72	49.18
Ours (SCIDER)	**67.00**	**72.11**	59.62	**46.69**	**58.06**	**36.15**	**56.47**

Table 4: The macro-averaged F1 score (%) and accuracy (ACC) (%) of our model and previous works on PDTB 2.0 and CoNLL16. Italics numbers indicate the reproduced results from (Chan et al., 2023b). Bold numbers correspond to the best results, whereas underlined numbers correspond to the second best.

Model	COMPARISON	CONTINGENCY	EXPANSION	TEMPORAL
NNMA (Liu and Li, 2016)	36.70	54.48	70.43	38.84
Gshare (Lan et al., 2017)	40.73	58.96	72.47	38.50
Bi-LSTM-DU (Dai and Huang, 2018)	46.79	57.09	70.41	45.61
ELMo-C&E (Dai and Huang, 2019)	45.34	51.80	68.50	45.93
RWP-CNN (Varia et al., 2019)	44.10	56.02	72.11	44.41
TransS (He et al., 2020)	47.98	55.62	69.37	38.94
BMGF-RoBERTa (Liu et al., 2020)	59.44	60.98	77.66	50.26
KANN (Guo et al., 2020)	43.92	57.67	73.45	36.33
CG-T5 (Jiang et al., 2021)	55.40	57.04	74.76	41.54
CVAE-IDRR (Dou et al., 2021)	55.72	63.39	**80.34**	44.01
Ours (SCIDER)	**63.92**	**66.67**	78.12	**61.02**

Table 5: Binary classification results on PDTB for the 4 top-level classes of our baseline and previous works in terms of macro-averaged F1 score (%).

- Our model achieves exceptional average performance compared with previous state-of-the-art models.
- Overall results demonstrate the potential by disentangling logical semantics from general semantics.

Visualization

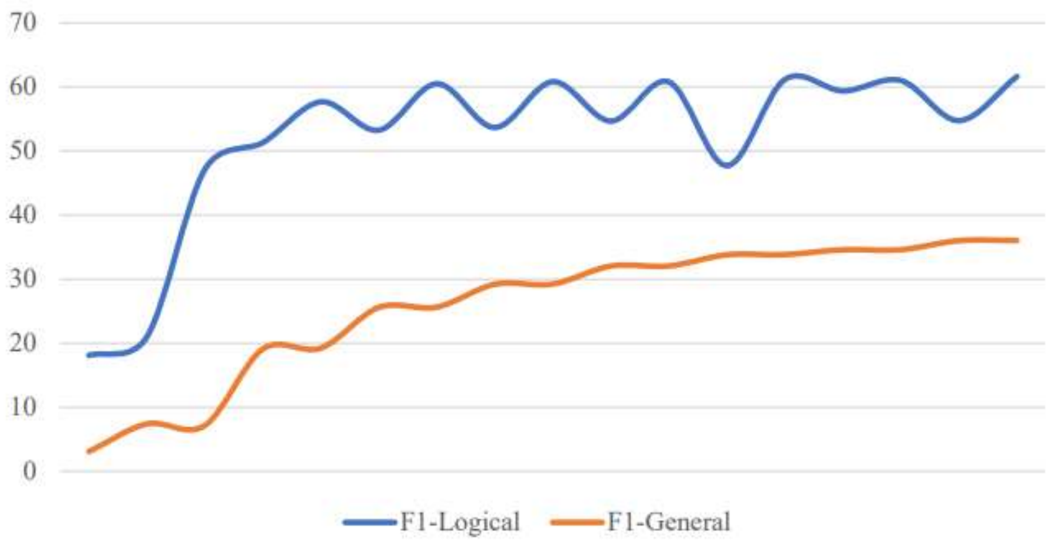


Figure 4: The trend of the F1 score during the training process. The results are obtained by conducting classification based on the embedding of $[logical]$ and $[general]$, respectively.

- The F1 scores of two semantics closely parallel in initial epochs.
- As training continued, a notable divergence emerged, which indicates a pronounced disentanglement between logical and general semantics.

Conclusion

- We argue that the representation learned by the PLMs, contains many aspects of semantics, e.g., general and logical semantics.
- A semantics confrontation method is proposed to disentangle the two semantics.
- Experimental results demonstrate that such disentanglement improves the performance in IDRR.

Thanks for your attention