

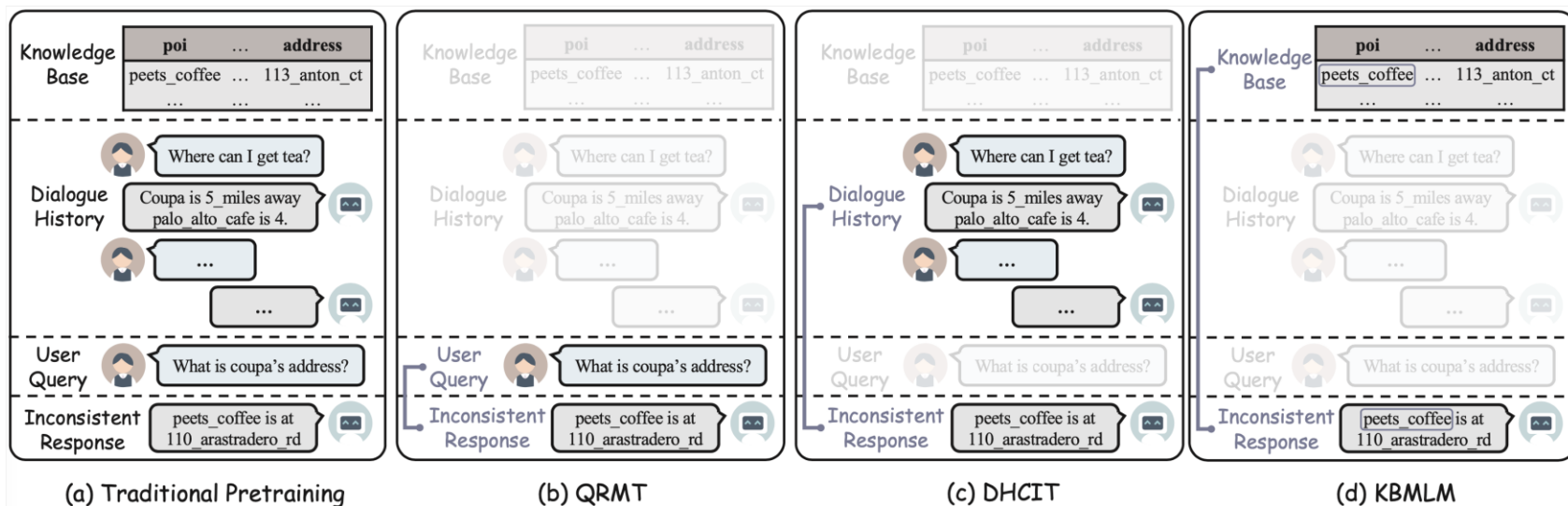
# MPFToD: A Modularized Pre-training Framework for Consistency Identification in Task-oriented Dialogue

**Libo QIN, Shijue HUANG, Qiguang CHEN, Qian LIU,  
Wanxiang CHE, Ruifeng XU**

Frontiers of Computer Science, DOI: [10.1007/s11704-024-3778-9](https://doi.org/10.1007/s11704-024-3778-9)

# Problems & Ideas

- Problems of consistency identification in a task-oriented dialogue (CI-ToD):
  - Existing pre-trained models are general-purpose, which lacks dialogue context understanding and KB reasoning ability.
  - Traditional CI-ToD pre-training approach that strictly requires a large number of multi-turn KB-grounded dialogues.
- Ideas: Decouple the pre-training into three sub-modules to make full use of different kinds of KB-free data resources.

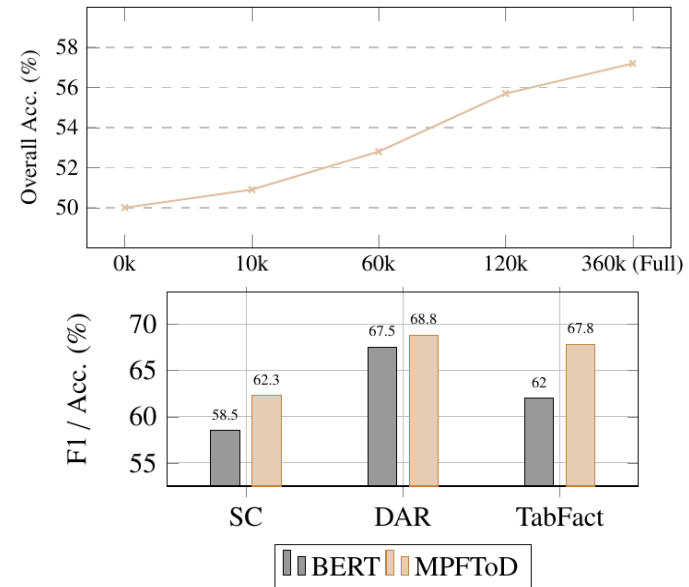


Different pre-training paradigms. (a) Traditional Pre-training; (b) Query Response Matching (QRMT); (c) Dialogue History Consistent Identification Task (DHCIT); (d) KB Mask Language Modeling (KBMLM).

# Main Contributions

- Contributions:
  - Take the first attempt to explore a modularized pre-training paradigm for CI-ToD, which allows the model to utilize large amounts of KB-free dialogues corpus;
  - Devise three different pre-training tasks for CI-ToD, which enhances the dialogue context understanding ability and KB reasoning ability;
  - Experimental results on the CI-ToD benchmark show that our method achieves state-of-the-art performance (with 4.7% improvement).

Model	QI F1	HI F1	KBI F1	Overall Acc.
<i>Non Pre-trained Models</i>				
ESIM [2]	0.512	0.164	0.543	0.432
InferSent [3]	0.557	0.031	0.336	0.356
RE2 [4]	0.655	0.244	0.739	0.481
<i>Pre-trained Models</i>				
BERT [5]	0.691	0.555	0.740	0.500
RoBERTa [6]	0.715	0.472	0.715	0.500
XLNet [25]	0.725	0.487	0.736	0.509
Longformer [26]	0.717	0.500	0.710	0.497
BART [7]	0.744	0.510	0.761	0.513
TOD-BERT [27]	0.741	0.459	0.766	0.535
CGIM [8]	0.764	0.567	0.772	0.563
MPFToD	0.777	0.576	0.788	0.572
MPFToD + CGIM	<b>0.837*</b>	<b>0.605*</b>	<b>0.814*</b>	<b>0.610*</b>



Main results and analyses. Left: main results on CI-ToD benchmark; Right: the results of different pre-training data sizes and the generalization results on other tasks including sentiment classification, dialogue act recognition and table fact checking.