Dialog Action-Aware Transformer for Dialog Policy Learning

Huimin Wang[2]*, Wai-Chung Kwan[1]*, Kam-Fai Wong[1]

[1]The Chinese University of Hong Kong
[2]Jarvis Lab, Tencent

Introduction

Dialog policy learning (DPL) plays a crucial role in pipeline task-oriented dialog systems by determining the next abstracted system action.

Pre-trained language model (PLM) does not work well in DPL due to **misalignment** of pre-training tasks with non-natural language.

Problems of Pre-training Tasks for DPL

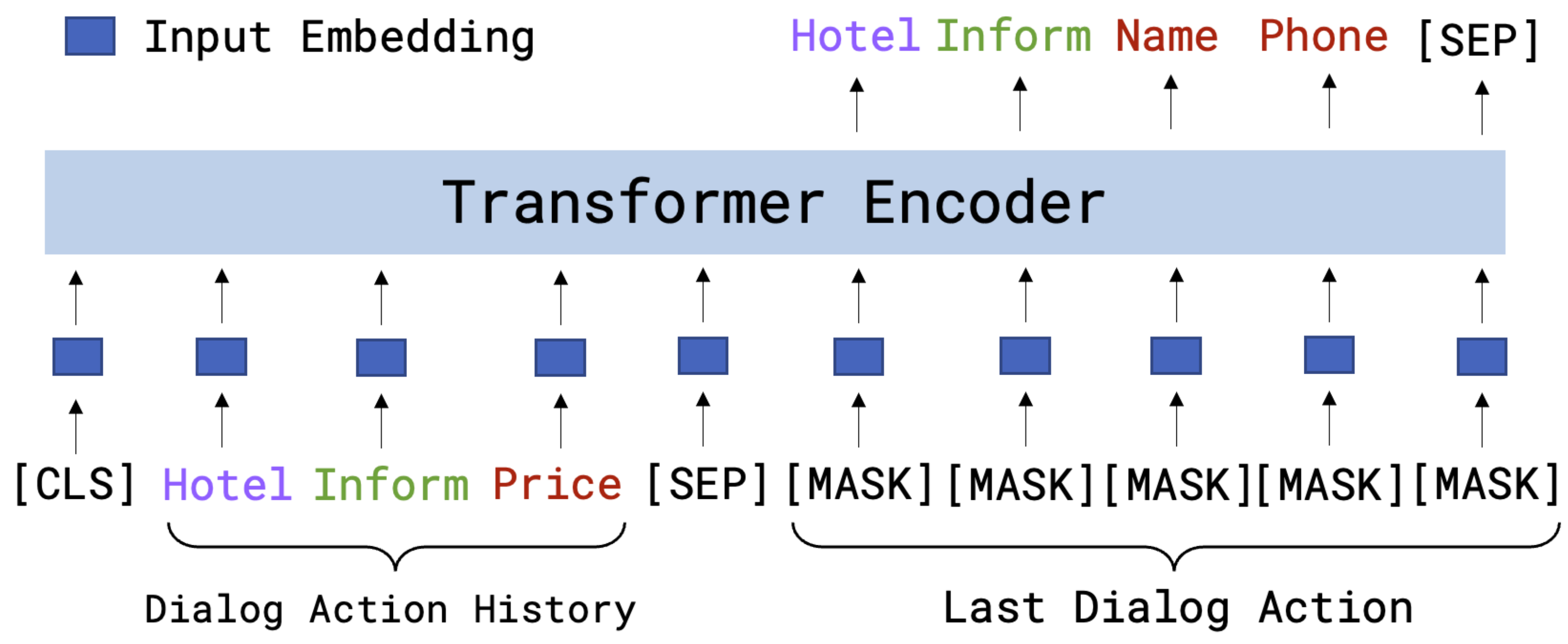
The **next sentence prediction** (NSP) task benefits understanding of natural language but not on structured non-natural language (e.g. dialog actions).

The **masked language modelling** (MLM) task fuses the content in both directions where the dialog agent is only allowed to access the left.

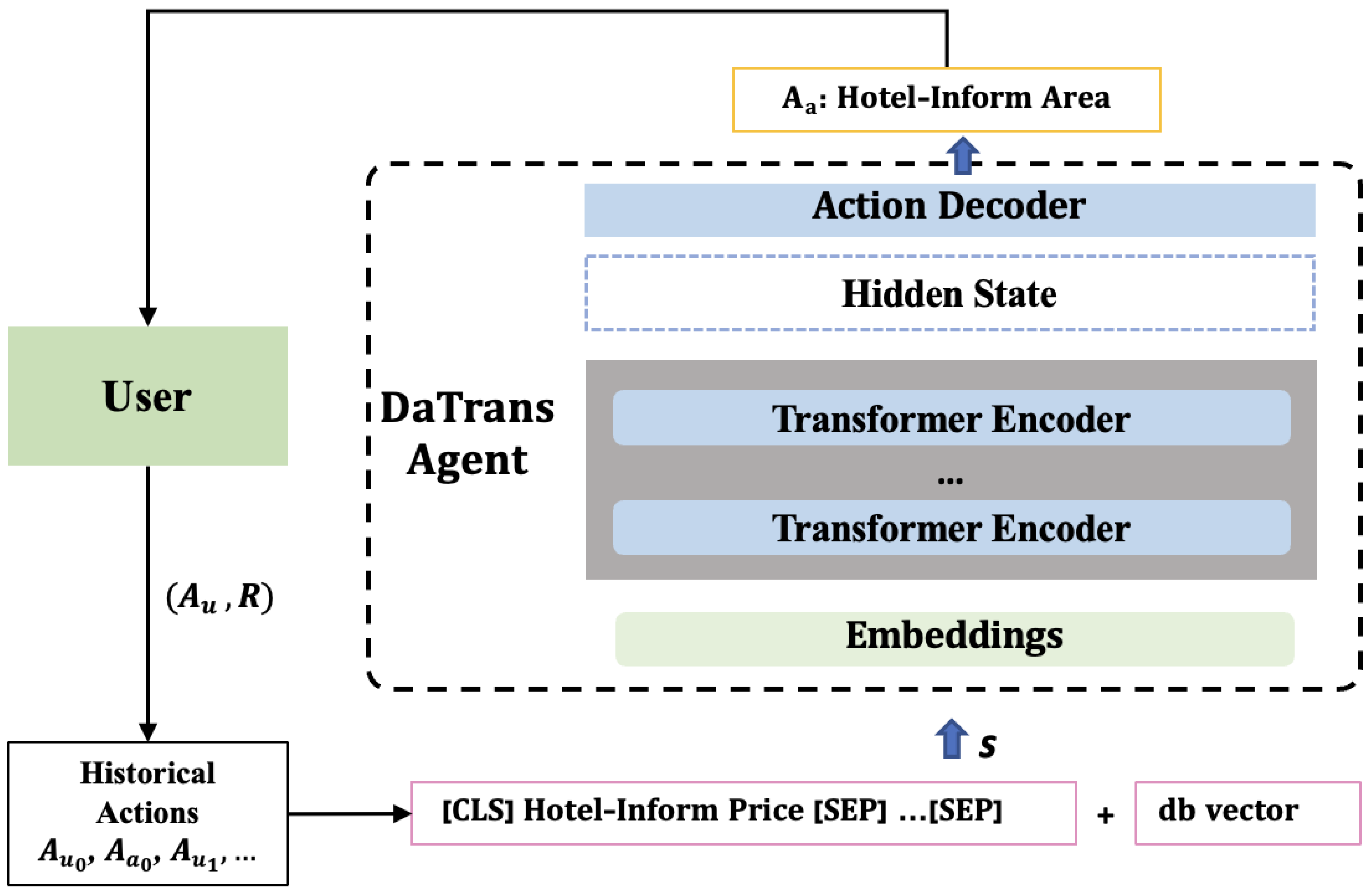
Method

We propose **D**ialog **A**ction-oriented **Trans**former (**DATrans**) for efficient DPL.

We propose a novel pre-training task **MLA**: predicting the **M**asked **L**ast dialog **A**ction.



We further fine-tune **DATrans** with **Deep Q-learning** using a user simulator.



Experiments

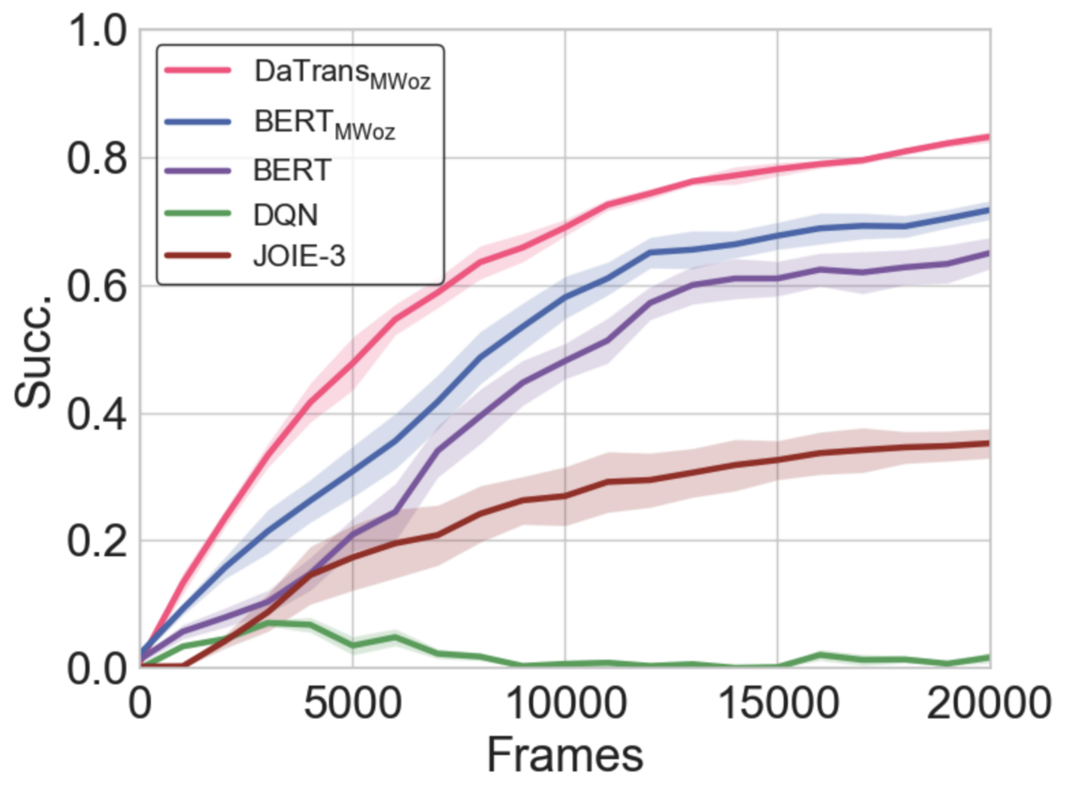
Datasets: MultiWoz and SGD. Two popular task-oriented dialogue dataset.

Baselines:

- **BERT$_{\text{MWoz}}$**: BERT pre-trained with MLM and NSP on MultiWoz and fine-tuned by Deep Q-learning.
- **BERT**: Fine-tuning pre-trained BERT on MultiWoz by Deep Q-learning.
- **DQN**: An MLP network optimized by Deep Q-learning.
- **JOIE**: Previous state-of-the-art using a collaborative multi-agent framework.

Main Results

Model	Succ.↑	Turn↓	Reward↑
DaTrans$_{\text{MWoz}}$	**0.84**	**10.21**	**27.35**
BERT$_{\text{MWoz}}$	0.72	12.14	14.21
BERT	0.64	14.75	-15.47
DQN	0.01	19.51	-53.66
JOIE-3	0.38	15.98	-21.42

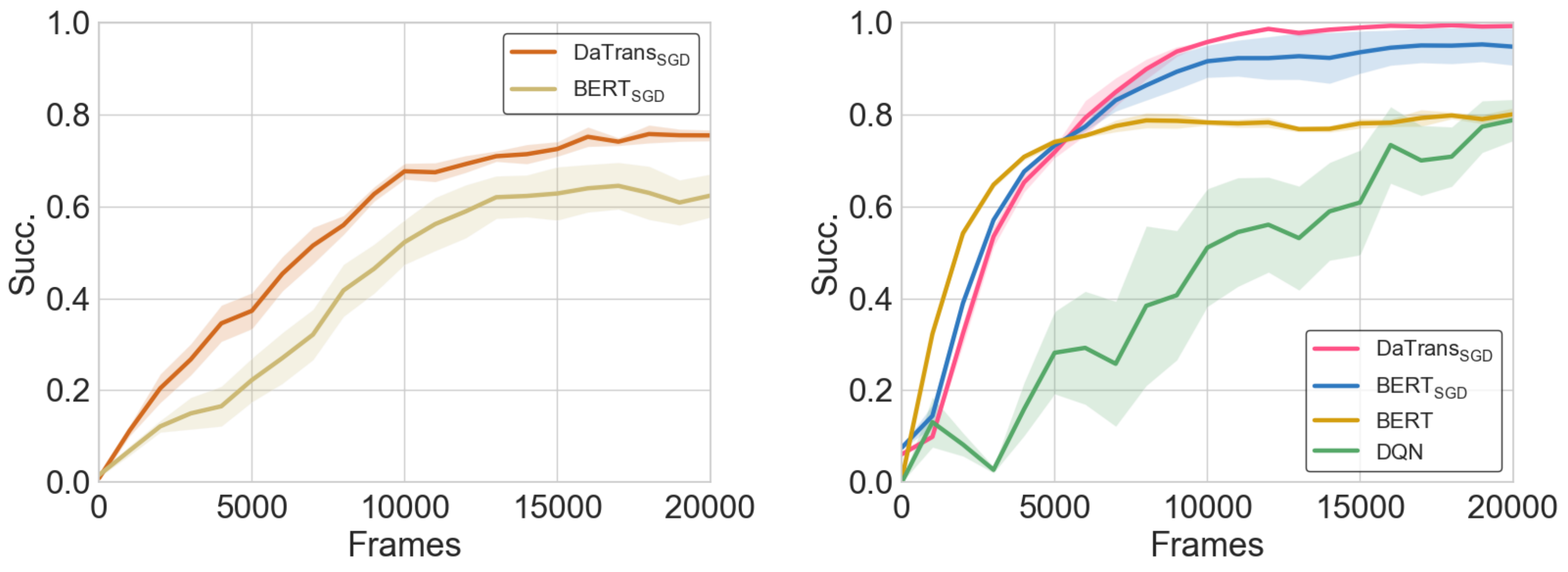


DaTrans$_{\text{MWoz}}$ > BERT$_{\text{MWoz}}$: MLA is better than NSP and MLM.

DaTrans$_{\text{MWoz}}$ > BERT: The pre-training misalignment can't be bridged by reinforcement learning alone.

Transfer Learning

Pre-trained on SGD, fine-tuned on MultiWoz.



Restaurant domain

All domains

DaTrans is robust to different pre-training corpus.

DaTrans adopts quickly to new domain.

Takeaway

- Language models pre-trained on large text corpus cannot be utilized in DPL.
- Pre-training on non-natural language corpus significantly enhances DPL (even with NSP and MLM tasks).
- Pre-training with MLA task outperforms NSP and MLM significantly in DPL (0.84 vs 0.72 success rate).
- Fine-tuning with RL is unable to bridge the misalignment gap caused by pre-training suboptimally.