Leveraging Transformer-based Pretrained Language model for Task-oriented dialogue system

MANISHA THAKKAR, NITIN N PISE
School of Computer Engineering and Technology
Dr Vishwanath Karad, MIT World Peace University, Pune, INDIA

Abstract: Task-oriented dialogue (TOD) systems are becoming increasingly popular due to their wide acceptability at personal and enterprise levels. These systems assist users to accomplish the intended task. Natural language input is given to such systems as text or speech. It is very complex to design TOD systems because it is difficult to maintain dialogue flow during the conversation. These neural systems require a lot of task-specific annotated data. To overcome the data scarcity of such systems recent advancements in Pretrained Language Models (PLMs) have shown promising results. In this paper, we studied the application of transformer-based PLMs to TOD systems tasks and compared their performances.

Keywords: Task-oriented dialogue system, End-to-End systems Pretrained Language models, text-to-text transfer Transformer (T5)

Received: March 30, 2022. Revised: November 22, 2022. Accepted: December 15, 2022. Published: February 3, 2023.

1. Introduction

Task-oriented dialogue (TOD) systems accomplish user goals by supporting interactions in the natural language and they are widely used in many applications, such as flight booking, and hotel reservations. It is crucial to keep track of dialogue flow and understand the change in dialogue. There are two main approaches to designing TOD systems. 1) Pipeline Approach 2) Neural Approach End to End (E2E) [1]

1) Pipeline approach: In this method, a traditional complex modular pipeline all the components of TOD systems namely, NLU (Natural Language Understanding) which performs intent recognition and domain identification from user utterance, DM (Dialogue Manager) which has two sub-components namely, DST (Dialogue State Tracking) for tracking user's belief state and Dialogue Policy (POL) for deciding which system action to take being in a given state and NLG (Natural Language Generation) for generating dialogue response in natural language are independently developed and trained, and modularly connected in a pipeline. The pipeline approach of designing a dialogue system has the following limitations.

Credit Assignment Problem: In the pipeline method, when the correct response is not generated, it is difficult to identify which module (NLU, DM, and NLG) is responsible for no response or bad response generation, or not achieving the task goal.

Process Interdependence: Any update in one dialogue system component requires retraining other components to sync. This takes additional time for each update.

These dialogue agents are ontology-based. New rules are required to be handwritten for each new task and domain, which is costly, time-consuming, and non–scalable. Ontology dependence is the main hurdle for the scalability of these agents. [2]

To overcome the limitations of the Pipeline approach, the E2E approach is a neural approach for designing TOD systems in that all the components of dialogue systems are trained together on the same dialogues without any assumptions of domain or dialogue state structure. This makes the system scalable to perform in new domains. These systems have shown promising performance in open-domain agents and are now popularly applied to TOD systems. But these neural models require huge task-specific annotated data for training such data is not always available and creating annotations is time-consuming and costly. The research community has witnessed remarkable progress of pre-training methods including BERT, GPT, and other transformer-based models in many of the NLP tasks such as language understanding, and language generation. [3]

In this paper, we represent the application of a text-to-text transformer, T5 for NLU, DST, and NLG components of the TOD system. We also compare the results with the application of BERT for NLU, GPT for NLG.

2. Literature Survey

TOD systems are becoming an increasingly popular and growing area of research. These agents take user input in the form of speech or text and help to achieve goals for example booking a taxi, reserving a table in a hotel, booking movie tickets, etc in the general domain. It is challenging for a dialogue system to understand the user goal and keep track of the user's dialogue flow and carry out an effective conversation when the user changes the domain or task frequently while having a conversation. The Natural Language Understanding (NLU) component of the dialogue system performs domain classification, intent classification, and slot filling from the user's utterance. [4]

For example when the user query is "I am looking for a cheap restaurant with Indian food.", then the domain here is "Restaurant", the intent is "Find Restaurant" and the slots are "price range" with value "cheap" and "food type" with slot value "Indian". NLU represents this query semantically as

inform (domain=' Restaurant', price range ="cheap", food type = "Indian"),

These slots are filed by tagging each word of the user

message. Whereas the DST module doesn't classify or tag user messages instead it tries to find a slot value for each name in an already defined slot list based on user dialogue history. In current Task-oriented dialogue agents, DM is designed with hand-crafted rules for a specific domain having its domain ontology which includes pre-defined (slot-value) pairs. Such systems are not scalable to multiple domains and tasks because they involve a lot of human effort and expertise.[5]

Pretrained Language models (PLMs) in recent years are becoming increasingly popular to address the problem of data scarcity. For example, GloVe pretrained word embedding which represent all instances of the same word with same vector representation, but it fails to extract meaning as context is not considered it also fails to handle out-of-vocabulary (OOV) problem decently. Using Pretrained models like ELMo and Generative Pre-trained Transformer (GPT), Bidirectional Encoder Representations from Transformers (BERT) and fine-tuning them on NLP tasks to achieve significant improvement over training on task-specific annotated data is in research trend[6]. T5 is an encoder-decoder model pre-trained on a multi-task mixture of unsupervised and supervised tasks and for which each task is converted into a text-to-text format. T5 works well on a variety of tasks. [7]

Below table describes the application of PLM for TOD tasks.

TABLE 1: APPLICATION OF PLMS FOR DIFFERENT TOD TASKS

Reference	Pretrained Model Applied	Task
[3], 2018	Pre-trained word representations ELMo	NLU
[6], 2019	bidirectional pre-training for language representations	NLU
[8], 2020	BERT	NLU
[9], 2021	BERT	NLU (Token-aware Contrastive Learning)
[10], 2019	GPT (decoder-only model)	NLG
[11], 2020	GPT-2	cascaded model, all TOD sub-tasks
[12], 2021	GPT-2	NLG
[13], 2021	GPT-2	TOD sub-tasks
[7], 2020	T5	NLG
[14], 2021	BART	NLG
[15], 2022	T5	pre-train model with all TOD-related tasks, dialogue context and the task specific prompt as input to generate the corresponding target text

Most of the researchers work considered single component of TOD task for application of PLM, language understanding and language generation are more focused.

3. Methodology

In this section, we first discuss the datasets and apply text-to-text transformer for NLU, DST and NLG task in Convlab -3 open source toolkit.[16]

Dataset: MULTIWOZ (MWOZ)

Multi domain multi turn Human – to –Human conversation dataset MULTIWOZ of size 10K is considered for experiment. This dataset is annotated for dialogue state, system dialogue act, user goals in different domains of trip information setting (Hotel, Train, Hospital, Taxi, Police, postcode, Restaurant).Fig. 2 describes the data from Restaurant domain from JSON file. [17]


```
{
    "address": "31 Newnham Road Newnham",
    "area": "west",
    "food": "indian",
    "id": "19254",
    "introduction": "indian house serve a variety of indian dishes to eat in or take away
     they also have a selection of english dishes on their menu",
    "location": [
      52.199012,
      0.113196
    ],
    "name": "india house",
    "phone": "01223461661",
    "postcode": "cb39ey",
    "pricerange": "expensive",
    "type": "restaurant"
  },
```


Fig. 2 Sample Data from Restaurant domain

There are 3, 406 single-domain dialogues and 7, 032 multi-domain dialogues which contains at least 2 to 5 domains. Almost 70% of dialogues have more than 10 turns which makes it complex with many real time scenarios. The ontology for all domains in MULTIWOZ data-set is described in Table 2. The upper script denotes the domain they belong to and the slots are divided into informable slots and requestable slots. In informable slot user specify constrain in the search. For eg. Area: south zone, Price: cheap etc. on the other hand in requestable slots user can ask for additional information like address, phone no etc.

TABLE 2: THE ONTOLOGY FOR ALL DOMAINS IN MULTIWOZ DATASET

Domain	*: universal , 1: restaurant, 2: hotel, 3: attraction, 4: taxi, 5: train, 6: hospital, 7: police,
act type	Inform* / request*/ select123 / recommend/123 / not found123 request booking info123 / offer booking1235 / inform booked1235 / decline booking1235 welcome*/greet*/ bye* / reqmore*
slots	Address* / postcode*/ phone*/ name[1234] / no of choices[1235] / area[123] / pricerange[123] / type[123] / internet[2] / parking[2] / stars[2] / open hours[3] / departure[45] destination[45] / leave after[45] / arrive by[45] / no of people[1235] / reference no.[1235] / trainID[5] / ticket price[5] / travel time[5] / department[7] / day[1235] / no of days[123]

Google's T5 Text-to-Text Transfer Transformer

Google's T5 treat every text processing problem as a "text-to-text" problem, i.e. taking text as input and generate new

text as output. This approach of text-to-text framework allows us to directly apply the same model, objective, training procedure, and decoding process to every task under consideration also there is flexibility to evaluate the performance on wide range of NLP problems such as Question-Answer, summarization, translation, classification etc. This model is trained on "Colossal Clean Crawled Corpus" (or C4 for short).

T5 v/s BERT and GPT

BERT is "encoder-only" model designed to produce a single prediction per input token or a single prediction for an entire input sequence. This makes them applicable for classification or span prediction tasks but not for generative tasks like translation or abstractive summarization. GPT uses only decoder block of transformer, and it is applied to text generation tasks only whereas T5 is encoder-decoder structure that achieved good results on both generative and classification tasks.

4. Experiment

We have applied generative T5 transformer for (1) Natural Language Understanding (2) dialogue state tracking; and (3) Natural Language Generation tasks for MultiWOZ 2.1 dataset.

NLU:

Input: Text entered by User, Dialogue history as Context
Output: Dialogue act [intent][domain]([slot][value],...); separated by ; for more values
Model used: t5-small fine-tuned on MultiWOZ 2.1. (From huggingface library)

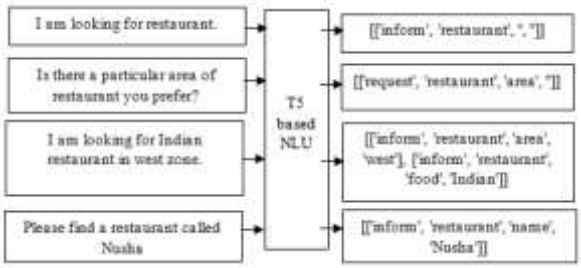

Fig. NLU with T5 at backend

DST:

Input: Dialogue history as Context
Output: State is in the form State is in the form of [domain]([slot][value],...); separated by ; for more values
Model used: t5-small fine-tuned on MultiWOZ 2.1. (From huggingface library)

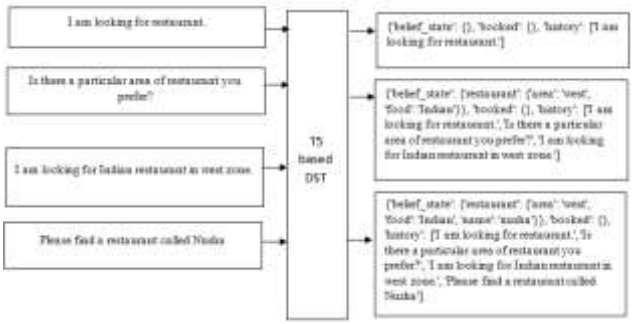

Fig. DST with T5 at backend

NLG:

Input: Dialogue Act and Context
Output: Language Generation
Model used: t5-small fine-tuned on MultiWOZ 2.1. (From huggingface library)

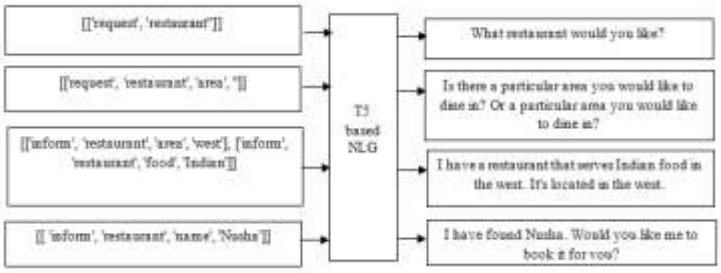

Fig. NLG with T5 at backend

Below table discuss the evaluation metrics of T5 and other State-of-the-art models

NLU Model	Acc	F1
BERTNLU	74.5	85.9
T5NLU	77.8	86.5

*Acc - Accuracy

DST Model	Joint Goal Acc	Slot F1
SetSUMBT	50.8	91.0
T5DST	53.1	91.9

NLG Model	Slot Error Rate ↓	BLEU
SC GPT	3.3	33.5
T5NLG	3.2	35.6

Comparision of T5 based NLU, ,DST and NLG with SOTA models.[7]

5. Conclusion and Future Work

Human dialogue is full of ambiguity which increases the complexity to understand user intent from respective domain which is essential to accomplish the task. PLMs are widely applied for various NLP Tasks due to significant improvement in performance to overcome data scarcity. By leveraging PLMs better model initialization can be achieved which helps in generalization along with downstream tasks and speedy convergence on target task.T5 transformer based tasks outperformed when compared with other SOTA models. In future, we plan to implement complete E2E dialogue system by leveraging PLMs for TOD.

References

[1] J. Gao, M. Galley, and L. Li, *Neural approaches to conversational AI*, vol. 13, no. 2–3. 2019.

[2] B. Liu and I. Lane, "End-to-End Learning of Task-Oriented Dialogs," pp. 67–73, 2018, doi: 10.18653/v1/n18-4010.

[3] M. E. Peters *et al.*, "Deep contextualized word representations," *NAACL HLT 2018 - 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 2227–2237, 2018, doi: 10.18653/v1/n18-1202.

[4] Y. Kim, "Convolutional neural networks for sentence classification," *EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, pp. 1746–1751, 2014, doi: 10.3115/v1/d14-1181.

[5] J. Pei, P. Ren, and M. de Rijke, "A Modular Task-oriented Dialogue System Using a Neural Mixture-of-Experts," 2019, [Online]. Available: http://arxiv.org/abs/1907.05346.

[6] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[7] C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, pp. 1–67, 2020.

[8] C. S. Wu, S. Hoi, R. Socher, and C. Xiong, "TOD-BERT: Pre-trained natural language understanding for task-oriented dialogue," *EMNLP 2020 - 2020 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, pp. 917–929, 2020, doi: 10.18653/v1/2020.emnlp-main.66.

[9] Y. Su *et al.*, "TaCL: Improving BERT Pre-training with Token-aware Contrastive Learning," *Find. Assoc. Comput. Linguist. NAACL 2022 - Find.*, no. 1, pp. 2497–2507, 2022, doi: 10.18653/v1/2022.findings-naacl.191.

[10] P. Budzianowski and I. Vuli, "Towards the Use of Pretrained Language Models for Task-Oriented Dialogue Systems," no. Wngt, pp. 15–22, 2019.

[11] E. Hosseini-Asl, B. McCann, C. S. Wu, S. Yavuz, and R. Socher, "A simple language model for task-oriented dialogue," *Adv. Neural Inf. Process. Syst.*, vol. 2020-December, no. NeurIPS, 2020.

[12] B. Peng, C. Li, J. Li, S. Shayandeh, L. Liden, and J. Gao, "Soloist: Building task bots at scale with transfer learning and machine teaching," *Trans. Assoc. Comput. Linguist.*, vol. 9, pp. 807–824, 2021, doi: 10.1162/tacl_a_00399.

[13] Y. Yang, Y. Li, and X. Quan, "UBAR: Towards Fully End-to-End Task-Oriented Dialog System with GPT-2," *35th AAAI Conf. Artif. Intell. AAAI 2021*, vol. 16, pp. 14230–14238, 2021, doi: 10.1609/aaai.v35i16.17674.

[14] S. Yang and Y. Liu, "Data-to-text Generation via Planning," *J. Phys. Conf. Ser.*, vol. 1827, no. 1, pp. 895–909, 2021, doi: 10.1088/1742-6596/1827/1/012190.

[15] Y. Su *et al.*, "Multi-Task Pre-Training for Plug-and-Play Task-Oriented Dialogue System," no. 1, pp. 4661–4676, 2022, doi: 10.18653/v1/2022.acl-long.319.

[16] Q. Zhu, C. Geishauser, H. L. Carel, X. Zhu, J. Gao, and M. Gaši, "ConvLab-3 : A Flexible Dialogue System Toolkit Based on a Unified Data Format."

[17] P. Budzianowski *et al.*, "MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling," pp. 5016–5026, 2019, doi: 10.18653/v1/d18-1547.