Conversational Topic Recommendation in Counseling and Psychotherapy with Decision Transformer and Large Language Models

Aylin Gunal
University of Michigan
Ann Arbor, MI
gunala@umich.edu

Baihan Lin
Icahn School of Medicine at Mount Sinai
New York, NY
baihan.lin@mssm.edu

Djallel Bouneffouf
IBM Research
Yorktown Heights, NY
djallel.bouneffouf@ibm.com

Abstract

Given the increasing demand for mental health assistance, artificial intelligence (AI), particularly large language models (LLMs), may be valuable for integration into automated clinical support systems. In this work, we leverage a decision transformer architecture for topic recommendation in counseling conversations between patients and mental health professionals. The architecture is utilized for offline reinforcement learning, and we extract states (dialogue turn embeddings), actions (conversation topics), and rewards (scores measuring the alignment between patient and therapist) from previous turns within a conversation to train a decision transformer model. We demonstrate an improvement over baseline reinforcement learning methods, and propose a novel system of utilizing our model's output as synthetic labels for fine-tuning a large language model for the same task. Although our implementation based on LLaMA-2 7B has mixed results, future work can undoubtedly build on the design.

1 Introduction

In recent years, there has been a notable uptick in the number of people seeking professional help for mental health concerns, but the available pool of mental health professionals remains small in comparison. To address this need, automated AI-based tools and methods for counseling have been explored and engineered, ranging from systems for training junior mental health counselors (Min et al., 2022; Demasi et al., 2019) to AI-in-the-loop chatbots (Sharma et al., 2022). With the dramatic rise in popularity and accessibility of large language models (LLMs), it's expected that LLMs will play a significant role in the intersection of computing and mental health research, as well.

In our prior work (Lin et al., 2023b), we introduced the SupervisorBot, a reinforcement learning (RL)-based topic recommendation system in counseling conversations. This proves to be a useful tool for clinicians during their psychotherapy sessions, where the system recommends what topics to discuss next given what has been discussed so far, as well as what works best in the past in terms of the patient outcomes. In this work, we improve upon this meaningful task by introducing the Decision Transformer (Chen et al., 2021), a transformer model designed for reinforcement learning (RL), into the recommendation pipeline, demonstrating better performance than other RL methods. We also explore the potential combination of Decision Transformer with LLMs, by generating labels for unseen transcript data using the pre-trained Decision Transformer model, and feeding the synthetically annotated data to fine-tuning a LLM. Our primary contribution is demonstrating that in the task of topic recommendation, Decision Transformer outperforms baseline RL methods; if such a system were to go through the process of user testing, the Decision Transformer—or models building on or improving Decision Transformer—can be utilized as the backbone for the recommendation module.

We first describe how we implement the preprocessing of the therapy conversation dataset, and how this is fed into the Decision Transformer model. We then describe how we use a portion of the dataset to train the Decision Transformer model, and that trained model's predicted labels are used as input to a large language model to train for the same task of topic recommendation.

2 Related Work

Decision Transformer was introduced as a transformer-based architecture to abstract the process of offline reinforcement learning, and has been used successfully in various NLP tasks including natural language understanding (Zhang et al., 2022; Bucker et al., 2023), navigating text-based games (Putterman et al., 2021), and generative language modeling (Memisevic et al., 2022). The Decision Transformer architecture has also been effectively

applied to the clinical domain to generate treatment recommendations based on patient history (Lee et al., 2023). In this work, we effectively apply the Decision Transformer architecture to the mental health domain in a dialogue recommendation task and improve on performances with older reinforcement learning methods.

The improvement of AI-in-the-loop tools to support humans in tasks has typically focused on human feedback, although more recent work has explored the potential for AI tools to improve themselves through a number of methods. (Saunders et al., 2022) demonstrates that a generative language model can improve its own outputs through fine-tuning on its own generations, and that the improvements are more significant as the model size increases. Generative models also have the advantage of being able to generate improvements to their own outputs (Zelikman et al., 2023).

In addition to models improving themselves, ensemble methods in which one model serves some intermediary purpose within the pipeline—e.g. data generation or filtering for input to another model—can be used for conversational modeling tasks as well (Huang et al., 2023). (Stiennon et al., 2020) uses an intermediary model's output as a reward function for another model, outperforming sole supervised learning from the source dataset. In this work, we explore a potential pipeline in which one model's output is used as synthetic data to train a language model for the task of topic recommendation. We consider the idea of AI supplementing a typical reinforcement learning with human feedback (RLHF) process by experimenting with how AI may be able to augment feedback, which can have significant implications given the lack of publicly available mental health dialogue data, let alone annotated data.

3 Architecture

In the following sections, we describe the architecture of our system in detail (see Fig. 1).

3.1 Decision Transformer

We re-implement the recommendation system pipeline as described in our original paper, (Lin et al., 2023b). This system is designed to provide real-time feedback in the form of next-topic recommendation for mental health counselors in session with patients, using reinforcement learning methods to learn and to recommend the next topic (the *action* taken by the counselor) to move on from the current segment of dialogue (the current *state*). *Rewards* are calculated using working alliance inventory (WAI) (Horvath and Greenberg, 1989), a score from a survey of questions to determine how aligned a counselor is with their patient within a session. WAI is determined by computing similarity between inventory items and segments of dialogue (Lin et al., 2023a), and inventory items fall under three different categories: Task, Bond, and Goal. We include an aggregate WAI score, referred to as Full.

The original SupervisorBot paper evaluates the system's performance on three baseline RL algorithms: DDPG (Lillicrap et al., 2015), TD3 (Fujimoto et al., 2018), and BCQ (Fujimoto et al., 2019).

We use the Alex Street dataset [1], a dataset composed of counseling session transcripts for patients suffering from depression, anxiety, suicidal thoughts, and schizophrenia. The Alex Street dataset is preprocessed and segmented into turn-pairs, which are then embedded using Word2Vec (Mikolov et al., 2013). We use embedded topic modeling (Dieng et al., 2020) to extract 8 topics from the corpus—as determined optimal by the motivating paper—and label each turn-pair with the topic it best represents. The WAI scores are computed for each turn-pair. As the original system design is done, turn-pair embeddings represent states, topic labels represent actions, and associated WAI scores represent rewards. These items are fed as input into the Decision Transformer in the form of tuples of (r_t, s_t, a_t).

We defer to the original Decision Transformer paper for architecture details. Our model contains a single-head, 3-layer attention mechanism, and we use a context window of 20 for the baseline results. Pearson's correlation between the model predicted actions and real actions taken is used for evaluation for all experiments. We run experiments 5 times using a 95%/5% train-test split, and take the average result.

3.2 LLMs for Recommendation

An attractive property of LLMs is flexibility in usage; at their core they simply model language probability distributions, making their outputs malleable to various tasks in NLP. In this section, we

[1] https://alexanderstreet.com/products/counseling-and-psychotherapy-transcripts-series

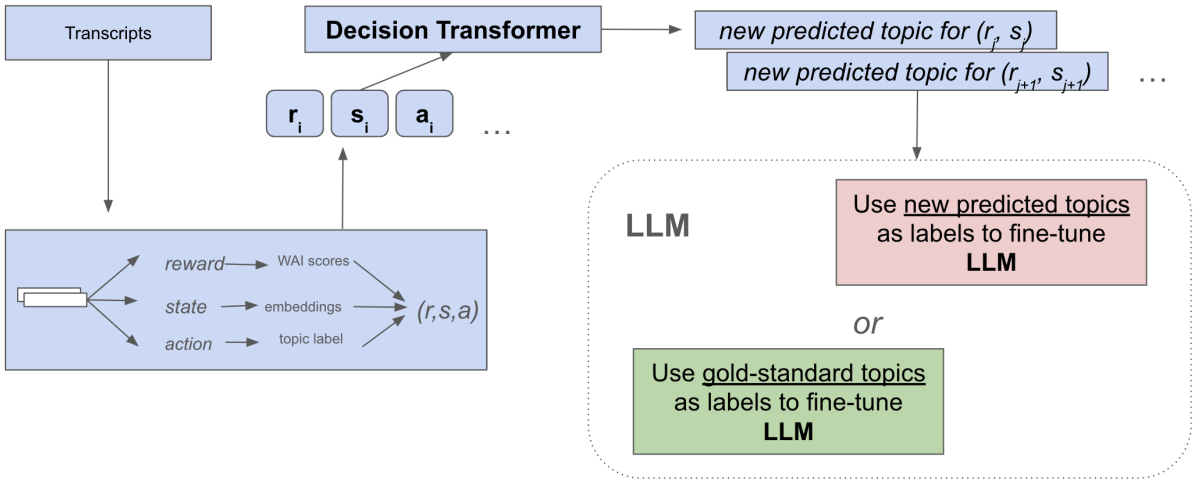

Figure 1: Architecture of the proposed LLM integration, demonstrating how both gold-standard labels from the dataset as well as synthetic annotations from Decision Transformer output can be used to fine-tune the LLM.

explore LLMs' ability for dialogue classification into labels of different psychiatric conditions to demonstrate their usefulness in various components of RLHF, similar to a diagnostic scenario in clinical setting as in (Lin et al., 2022, 2024). We primarily experiment with LLaMA-2 with 7B parameters (Touvron et al., 2023). We fine-tune LLaMA-2 model for sequence classification and test using the same preprocessing steps and train-test split described in Section 3.1. During the fine-tuning process, we do not use a validation dataset to avoid data leakage into the test set.

We experiment with treating Decision Transformer predictions as synthetic gold standard annotations for the LLM to learn from. We split the full dataset in a 40%/40%/20% split; the first 40% of the Alex Street dataset is used to train Decision Transformer, then Decision Transformer outputs predictions for another 40% of the dataset which the LLM is then fine-tuned with, and ultimately the LLM is evaluated on the final 20% of the dataset. Due to computational constraints, we apply low-rank adaptation (LoRA) (Hu et al., 2021), a parameter efficient fine-tuning method, in order to optimize the fine-tuning process. The LoRA configuration includes an alpha of 16 and dropout rate of .05, and we fine-tune for 1 epoch. We target the LLaMA-2 model's attention layers during training and save the final layer's weights to avoid those scores being randomly initialized for inference. To provide some baseline for comparison, we additionally fine-tune the LLaMA-2 model on the original gold-standard labels.

4 Evaluation

4.1 Results

The results of the Decision Transformer on a 95%/5% train-test split, reflecting the set-up of the original SupervisorBot paper, are provided in Table 1. We reproduce results for the other RL methods in the original paper for performance on the full-scale rewards; Decision Transformer outperforms these baselines as noted in Table 3. We note that Decision Transformer specifically performs best for all reward scales when trained on the full dataset; among individual diseases, the model performs best on the task, bond, and goal scales for anxiety.

We additionally evaluate whether or not the 20-timestep context is necessary for good performance from the Decision Transformer model, and these results are provided in Table 2. We note that 15 time-steps is optimal for a majority of the reward sclaes, suggesting that the Decision Transformer is better able to make decisions provided a briefer learning history. An advantage of utilizing a transformer-based model for this task is that we are able to investigate its internal structure to understand specifically which historical features—including which time-steps—are significant for inference.

Additionally, we note that the LLaMA-2 model trained on the gold-standard data does not necessarily outperform the Decision Transformer for all reward scales as indicated in Table 4, indicating that the off-the-shelf language model may not be conducive for a reinforcement learning task. LLaMA-2 trained on the Decision Transformer output directly

Decision Transformer

	Depression	Anxiety	Schizophrenia	Suicidal	All
Full	.176	.233	.246	.213	.361
Task	**.291**	**.320**	.247	.231	.323
Bond	.270	.314	.231	**.239**	.335
Goal	**.291**	.313	**.249**	.229	**.375**

Table 1: Results of Decision Transformer on topic recommendation task, using previous 20 turn-pairs as input. Best results per data subset are in bold.

	Context Lengths			
Rewards	5	10	15	20
Full	0.346	0.345	**0.403**	0.361
Bond	0.284	0.343	**0.359**	0.335
Task	0.272	0.298	**0.342**	0.322
Goal	0.278	0.339	0.348	**0.375**

Table 2: Decision Transformer model performance trained on varying context lengths. Best results per reward scale are in bold.

	DDPG	**BCQ**	**TD3**
Full	.264 (-.97)	.170 (-1.91)	.286 (-.75)

Table 3: Baseline RL performance on full-scale rewards on the full dataset, with a comparison to DT performance (the negative values suggest that these baselines perform worse than our proposed DT-based recommendation model, which improves upon them).

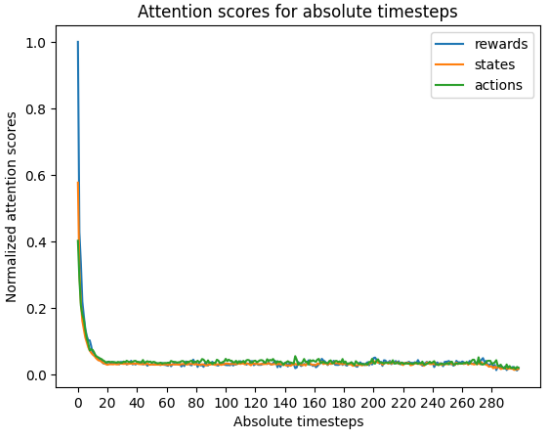

Figure 2: Normalized attention scores associated with absolute timesteps, *without* padded sequences.

also does not perform particularly well; future work may include modifying the way in which the Decision Transformer synthetic labels are used by a language model. It's possible that prompting the language model may yield better results than treating it as a sequence classifier.

4.2 Additional Analysis for Interpretability

We extract the final layer of attention weights from the Decision Transformer models trained on the four reward scales for the three types of inputs: returns, states, and actions. We observe both the attention weights for the individual input types as well as the aggregated and averaged set of weights

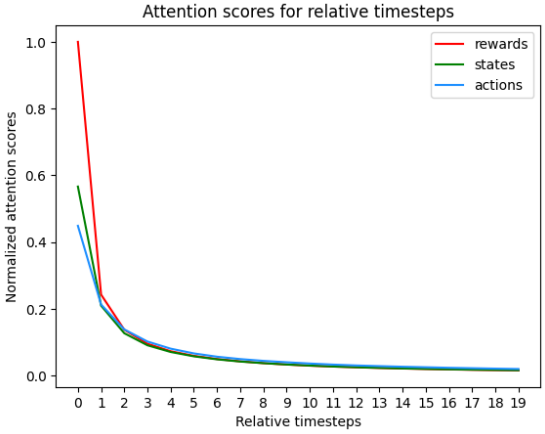

Figure 3: Normalized attention scores associated with relative timesteps.

	Full	Task	Bond	Goal
LLaMA-2 7B + DT	.148	.118	.158	.115
LLaMA-2 7B + Gold	.371	.259	.315	.332

Table 4: Results of fine-tuning LLaMA-2 7B on DT output and gold-standard labels.

aross all input types. Due to the auto-regressive nature of Decision Transformer, attention weights are assigned to the *timesteps* prior to the recommendation made at a given timestep.

We provide visual analyses of attention scores through normalized aggregate attention scores per timestep for absolute timestep values (Fig. 2) as well as relative timestep values (Fig. 3). We note that the model refines its attention to generally focus on items in earlier positions in given input sequence, both in the case for absolute and relative timesteps. These results, in tandem with the generally higher performances of the model on 15 previous timesteps rather than 20 timesteps, indicate that potentially there is a beginning index to the current context that can be key for the model's inference ability. Future work may include adjusting the context window dynamically, both for training and inference.

5 Limitations

Due to limitations of computational resources, experimentation with fine-tuning LLMs is restricted by model size. Future work can build on this work by applying similar experiments on increasing model sizes or non-quantized versions of models, effectively demonstrating (positively or negatively) that performance scales with model size.

6 Ethical Considerations

When implementing a topic recommendation system in counseling contexts, ethical considerations are important due to the sensitive nature of digital mental health discussions, as discussed in (Lin, 2022). One of the primary concerns is the potential limitation imposed by a static set of discussion topics. While such a system can streamline the counseling process, it risks limiting the creativity and flexibility of counselors, particularly those in training, and in the long term, inhibit consideration of their own perspectives on how to continue the conversation. This could inadvertently restrict their ability to tailor sessions according to the unique needs of each patient.

This is particularly relevant since the topics pulled are from one specific dataset that covers only four mental health conditions. The training dataset, derived from this limited number of mental health conditions, might not be representative of the broader population or other conditions. This limitation can lead to biased recommendations if not carefully managed. To mitigate this, it is essential to consider a more dynamic approach where the set of topics can evolve based on ongoing input from practicing counselors and feedback from therapy sessions. This adaptation would help in maintaining the relevance and sensitivity of the recommendations to diverse patient needs. In deployment, we can also imagine that topics are dynamically chosen, or chosen using human feedback; for example, perhaps before the system is put into use, counselors can input their own topics.

In addition to dataset limitations, the calculation of rewards, based on the Working Alliance Inventory (WAI), while rooted in established psychological theory, may benefit from enhancements through reinforcement learning with human feedback (RLHF). Incorporating direct input from users could refine the understanding and alignment of counselor and patient goals, improving the system's effectiveness and ethical alignment.

7 Conclusion

In this study, we introduced a Decision-Transformer-based recommendation system which outperforms baseline RL-based methods in counseling topic recommendation, indicating that transformer-based methods may have better performance in general when it comes to modeling conversation direction and alignment. We additionally find that the model performs best for certain reward scales on shorter input sequences, indicating that some exploration of optimal sequence length can be an avenue for future work. Through additional analysis of the attention scores, we additionally find that the model pays more attention to items earlier on in the input sequence.

References

Arthur Bucker, Luis Figueredo, Sami Haddadin, Ashish Kapoor, Shuang Ma, Sai Vemprala, and Rogerio Bonatti. 2023. Latte: Language trajectory transformer. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7287–7294. IEEE.

Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Michael Laskin, P. Abbeel, A. Srinivas, and Igor Mordatch. 2021. Decision transformer: Reinforcement learning via sequence modeling. In *Neural Information Processing Systems*.

Orianna Demasi, Marti A. Hearst, and Benjamin Recht. 2019. Towards augmenting crisis counselor training by improving message retrieval. *Proceedings of the*

Sixth Workshop on Computational Linguistics and Clinical Psychology.

Adji B Dieng, Francisco JR Ruiz, and David M Blei. 2020. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453.

Scott Fujimoto, Herke Hoof, and David Meger. 2018. Addressing function approximation error in actor-critic methods. In *International conference on machine learning*, pages 1587–1596. PMLR.

Scott Fujimoto, David Meger, and Doina Precup. 2019. Off-policy deep reinforcement learning without exploration. In *International conference on machine learning*, pages 2052–2062. PMLR.

Adam O Horvath and Leslie S. Greenberg. 1989. Development and validation of the working alliance inventory. *Journal of Counseling Psychology*, 36:223–233.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Xin Huang, Kye Min Tan, Richeng Duan, and Bowei Zou. 2023. Ensemble method via ranking model for conversational modeling with subjective knowledge. In *Proceedings of The Eleventh Dialog System Technology Challenge*, pages 177–184.

Seunghyun Lee, Da Young Lee, Sujeong Im, Nan Hee Kim, and Sung-Min Park. 2023. Clinical decision transformer: intended treatment recommendation through goal prompting. *arXiv preprint arXiv:2302.00612*.

Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2015. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Baihan Lin. 2022. Computational inference in cognitive science: Operational, societal and ethical considerations. *arXiv preprint arXiv:2210.13526*.

Baihan Lin, Djallel Bouneffouf, Yulia Landa, Rachel Jespersen, Cheryl Corcoran, and Guillermo Cecchi. 2024. Compass: Computational mapping of patient-therapist alliance strategies with language modeling. *arXiv preprint arXiv:2402.14701*.

Baihan Lin, Guillermo Cecchi, and Djallel Bouneffouf. 2022. Working alliance transformer for psychotherapy dialogue classification. *arXiv preprint arXiv:2210.15603*.

Baihan Lin, Guillermo Cecchi, and Djallel Bouneffouf. 2023a. Deep annotation of therapeutic working alliance in psychotherapy. In *International workshop on health intelligence*, pages 193–207. Springer.

Baihan Lin, Guillermo Cecchi, and Djallel Bouneffouf. 2023b. Supervisorbot: Nlp-annotated real-time recommendations of psychotherapy treatment strategies with deep reinforcement learning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 7149–7153.

Roland Memisevic, Sunny Panchal, and Mingu Lee. 2022. Decision making as language generation. In *NeurIPS 2022 Foundation Models for Decision Making Workshop*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Neural Information Processing Systems*.

Do June Min, Verónica Pérez-Rosas, Kenneth Resnicow, and Rada Mihalcea. 2022. Pair: Prompt-aware margin ranking for counselor reflection scoring in motivational interviewing. In *Conference on Empirical Methods in Natural Language Processing*.

Aaron L Putterman, Kevin Lu, Igor Mordatch, and Pieter Abbeel. 2021. Pretraining for language conditioned imitation with transformers.

William Saunders, Catherine Yeh, Jeff Wu, Steven Bills, Ouyang Long, Jonathan Ward, and Jan Leike. 2022. Self-critiquing models for assisting human evaluators. *ArXiv*, abs/2206.05802.

Ashish Sharma, Inna Wanyin Lin, Adam S. Miner, David C. Atkins, and Tim Althoff. 2022. Human–ai collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence*, 5:46–57.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. 2020. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems*, 33:3008–3021.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

E. Zelikman, Eliana Lorch, Lester Mackey, and Adam Tauman Kalai. 2023. Self-taught optimizer (stop): Recursively self-improving code generation. *ArXiv*, abs/2310.02304.

Ziqi Zhang, Yile Wang, Yue Zhang, and Donglin Wang. 2022. Can offline reinforcement learning help natural language understanding? *arXiv preprint arXiv:2212.03864*.