Clinically-Dependent Fully Automatic Treatment Planning System via Reinforcement Learning

Purpose/Objective(s):

Although fully automated treatment planning system (TPS) has several advantages, such as the ability to treat more patients and optimize treatments, clinics have not adopted it because there is a wide variation of practices between them. Additionally, dosimetrists make complex compromises while manually optimizing with a TPS, which is too complex to be captured by a metric, or not computable in a reasonable time. Here, we propose a solution adaptable to each clinic's practices: a reinforcement learning (RL) agent trained to mimic human dosimetrists' optimization on a cohort of previously treated patients. We hypothesize that by training one agent for each clinic, we ensure that guidelines specific to each of them are followed.

Materials/Methods:

RL agents adapt actions to situations where there are interactions with an environment. RL only needs a reward after performing an action. In the case of dose optimization, adjusting the weights of the constraints are the actions. The key is to find a way of rewarding the agent when making good decisions (actions) versus bad ones. Current RL methods in dosimetry struggle to mimic human-optimized plans. We propose a new reward system based on the dose distribution of past clinical cases, via calculating the DVHs differences between the agent dose, and the database dose. This would better guide the RL agent towards clinically-acceptable treatment plans. Most importantly, it also allows the optimization to fit each center's internal standard practices and guidelines.

Results:

We successfully trained agents to mimic the dose type of several clinics. We generated a cohort of 50 patients to train them and manually optimized the dose according to three guidelines. We then generated 20 other patients for testing purposes. The table shows the average difference between clinical doses and the ones optimized by our RL agents. Agents specializing in one type of guideline

managed to mimic it, but performed poorly on others. Thus, for a clinically helpful, fully-automated TPS, one RL agent should be trained for each clinical guideline.

Conclusion:

By leveraging past clinical dose data, we have demonstrated the feasibility of training RL agents to mimic human-optimized radiotherapy plans following specific clinical guidelines. The results show that agents trained on specific clinic guidelines perform better in mimicking those guidelines than a single, general-purpose agent. This finding supports our hypothesis that a fully automatic TPS tailored to each clinic's practices is achievable. Future work could involve expanding the patient cohort to non-phantom cases, including modalities other than prostate cases, and real-world testing with human oversight to ensure the safety and efficacy of the RL-based TPS. Our research could pave the way for developing clinically-dependent automated TPS.

Table of average DVHs distances on test cases:

	To Clinic A	*To Clinic B*	*To Clinic C*
RL on clinic A	**1.6**	2.2	5.1
RL on clinic B	2.3	**1.3**	2.3
RL on clinic C	2.7	2.5	**1.6**