Abstract

**Title**

Clinically-Dependent Fully Automatic Treatment Planning System

*… with DVHs-based dose distance RL?*

*Cite? => if allowed by the platform*

Purpose/Objectives

Modern radiation therapy relies on precise techniques like IMRT or V-MAT. Doctors solve complex optimization problems to deliver the best dose: minimizing harm to healthy tissue while sufficiently irradiating to kill the cancer. This process is subjective and lacks a standardized approach. As a result, treatment planning heavily depends on doctor expertise, leading to variability between clinical centers.

A fully automatic Treatment Planning System (TPS) would allow hospitals to treat more patients. It also allows online re-optimization. This would reduce the necessary margins, and reduce the toxicity of radiotherapy treatments.

The main reason why a fully automatic TPS is not adopted by clinical centers is that all methods that were tried so far are single-metric based. If it is meta-optimization, researchers need to define a metric to minimize/maximize. If it is reinforcement learning (RL), the reward is also based on the evaluation of a plan via a metric.

We believe that the complexity of the compromises made by dosimetrists while optimizing manually with a TPS is too complex to be captured by a metric (or not computable in reasonable time). ***More importantly, practices are different among different centers, and a dose that is acceptable for one clinic may not be in another clinic.*** This suggests the creation of one metric per clinic for measuring the quality of an optimized dose. This is highly impractical.

Hence, we propose a solution that is adaptable to each clinic’s own practices: a RL agent that is trained to mimic the optimization made by human dosimetrists on a cohort of previously treated patients. Our hypothesis is that by training one agent for each hospital, we could ensure that guidelines specific to each of them are followed.

Materials/Methods

Reinforcement learning is adapted to situations where there are interactions with an environment, through actions. RL agents learn by experimenting; we therefore need some previously treated patients to train. As training data, RL agents solely require a reward after taking some actions. RL agents then learn what action leads to the best cumulative reward.

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 taking 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 approach aims to better guide the RL agent towards clinically-acceptable treatment plans. ***But most importantly, it also allows the optimization to fit each center's internal common practices and guidelines.***

We separately trained five deep learning agents to optimize doses such that the final dose is as close as possible to each of our five guidelines. RL agents are taking the current DVHs of the optimized dose, and the weights values used, and predict the weights changes to perform.

Results

We managed to train agents to mimic the dose type of several clinics. We also trained one general agent, trying to mimic all five guidelines at the same time.

As this is ongoing research, we have generated a cohort of 100 patients for training, and manually optimized the dose according to 5 guidelines. We then generated 50 other patients for testing purposes. [Results presented are only about test patients, not used for training.]

The average difference between clinical doses and ones optimized by our RL agents is summarized in the table. We observe that agents specializing on one type of guideline manage to mimic this guideline, but perform poorly on other guidelines.

This comforts the idea that for a clinically useful fully automatic TPS, one should train one RL agent for each clinic guideline.

Conclusions

We have demonstrated the feasibility of training RL agents to mimic human-optimized radiotherapy plans following specific clinical guidelines, by leveraging past clinical doses data. The results show that agents trained on specific clinic guidelines achieve superior performance in mimicking those guidelines compared to 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 will involve:

* Expanding the patient cohort to non-phantom cases.
* Try on modalities other than prostate cases.
* Real-world testing with human oversight to ensure the safety and efficacy of the RL-based TPS.

We believe this research could pave the way for the development of clinically-dependent automatic TPS.