Clinical Knowledge Based Reinforcement Learning for Radiotherapy Treatment Planning System

Paul Dubois[1,2], Nikos Paragios[1], and Paul-Henry Cournède[2]

[1]TheraPanacea, Paris, France
[2]MICS, CentraleSupélec, Université Paris-Saclay, Paris, France

Abstract Achieving optimal dose distribution in radiation therapy planning is a complex task, with contradicting goals. Yet, this step is crucial with profound implications for patient treatment and toxicity management.
The absence of universally agreed-upon constraints prioritization in radiation therapy planning complicates the definition of an optimal plan, requiring a delicate balance between multiple objectives. This balanced usually ends up being done manually.
The optimization process is further hindered by complex mathematical aspects, involving non-convex multi-objective inverse problems with a vast solution space. Expert bias introduces variability in clinical practice, as treatment planning is shaped by the preferences and expertise of radiation oncologists and medical physicists.
To surmount these challenges, we propose a first step towards a fully automated approach, using an innovative deep learning method that allows automatic navigation towards acceptable solution. We successfully trained an agent evaluating actions of a human dosimetrist, in order to reach a plan similar to past history. As this is very new and ongoing research, we generated synthetic phantom patients and associated trust-able clinical dose. In future work, we hope to be able to apply this technique to real cases.

1 Introduction

In contemporary radiation therapy, photon intensity modulated radiation therapy (IMRT) stands as a pivotal technique utilized to attain precise and conformal dose distributions within target volumes. This achievement owes its realization chiefly to the advent of the multileaf collimator (MLC)[add citation].
Radiation therapy is now a reliable treatment for oncology [add citation]. Despite this consensus, the way to deliver radiotherapy for its best result remain very dependent upon doctors. Moreover, it appears that there is a large variability across centers[add citation?].
To achieve the best treatment, doctors need to solve a complex inverse mathematical optimization problem with multiples trade-offs[add citation]. There is a lack of standardized prioritization of constraints make the optimization a real challenge[add citation]. The standard procedure nowadays is to manually guide computer optimization: dosimetrists manually update the settings of an optimizing software (so called Treatment Planning System)[add citation].
There has been many tries to create a metric that quantify the quality of a treatment plan: Normal tissue complication probabilities (NTCP), Target coverage, Conformity index, Heterogeneity index (non-exhaustive list)[add citations]. However, none of them has been able to satisfy all radio-oncologists, and the only reliable way of assessing a plan for doctors is to check out the dose-volume histograms (DVHs) them-self.
As a result, Pareto surface exploration are doomed to failure due to the lack of impartial quantitative measurement for a particular plan[add citation]. Other meta-optimization techniques are similarly bounded, for the same reason[add citation]. An extra challenge to attend for those is the fact that not all cases have the same "difficulty". Hence, for an "easy" case, doctors will require an excellent dose (in terms of the metrics mentioned above), while they can be more permissive for "harder" cases. This make the acceptability of a plan hard to define in general.
Reinforcement learning is a machine learning paradigm concerned with training agents to make sequential decisions in dynamic environments. Through a process of trial and error guided by rewards or penalties, agents learn to optimize their actions to achieve long-term objectives. It appears that the decisions taken by dosimetrists when performing the optimization of a treatment can be formalized as a reinforcement learning problem. Moreover, dosimetrists can guide the TPS towards an acceptable plan, but they usually struggle explaining their decision while interacting with the TPS. This suggest the use of deep reinforcement learning, over expert base methods.

2 Materials and Methods

We introduce a new paradigm in reinforcement learning (RL), based on the evaluation of states, rather than the reward.

2.1 Reinforcement Learning Reward

In classical RL, we want $V(S_t) = R_t + \gamma V(S_{t+1})$ (so the update is $V(S_t) \leftarrow (1-\alpha)V(S_t) + \alpha\,[R_{t+1} + \gamma V(S_{t+1})]$). In the context of dose optimization, the reward R_t is defined as $R_t = \mathscr{E}(S_{t+1}) - \mathscr{E}(S_t)$. Where $\mathscr{E}$ is a function that evaluates the quality of a state (such that higher is better; if lower is better, then swap s_t and S_{t+1}).
The evaluation $\mathscr{E}$ can be one, or a mixture of the metrics mentions in introduction (Section 1) [add citation]. This setup may leverage knowledge about which actions to perform, instead of guessing randomly as a meta optimizer would do. We can hope to gain some computation time.

However, this technique is not using the plan used in past cases; it only needs the optimizer inputs (CT, structures contours). We propose to use the availability of past treatment plans, to better catch the complexity of decision taken by dosimetrists, and match better their expectations of a fully automatic treatment planning system.
As developed in previous work, we can derive a distance between doses plans [add citation]. If we consider the clinical dose of past cases (used for training) as the best achievable one, then we can evaluate a dose plan by computing its distance with the clinical dose plan.
Letting D_t be the dose associated with S_t, and D_C the clinical dose. We then define $\mathscr{E}(S_t) = \mathscr{D}(D_t, D_C)$. Since in that case, lower is better, we will define the reward as

$$R_t = \mathscr{E}(S_t) - \mathscr{E}(S_{t+1}) = \mathscr{D}(D_t, D_C) - \mathscr{D}(D_{t+1}, D_C).$$

This reward can be interpreted as the "distance gained to the clinical dose".

2.2 Reward-Free Reinforcement Learning

3 Results

4 Discussion

Appendix

Synthetic phantom patients

As this is very new and ongoing research, we generated synthetic phantom patients and associated trust-able clinical dose. In future work, we hope to be able to apply this technique to real cases.

Clinical dose

Optimization

Evaluation

References