Using deep bayesian neural networks for optimal treatment assignment in precision oncology. A contextual bandit problem.

Niklas Rindtorff, Nisarg Patel, Ming Yu Lu, HuaHua Zheng
Harvard Medical School

1 Introduction

The goal of precision medicine is providing the right treatment to the right patient at the right time. Oncology has been one of the central fields in modern precision medicine [1], with an increasing number of targeted compounds available, that are hypothesized to show activity in a very specific subset of patients (among the most popular examples being Imatinib for BCR-ABL positive CML).

Despite these early successes, assigning patients to adequate treatments remains a challenge until today. The current best practice in precision oncology is to base the treatment decision on published and frequently used therapeutic protocols that consider the patient's clinical characteristics and cancer genetics. For example, based on the status of a single biomarker, such as a BRAF V600E mutation, a treatment decision can be made [2].

These decision rules, or therapeutic protocols, are constantly evolving and used to link patients to optimal potential outcomes. Most therapeutic protocols in precision medicine are the result of biological reasoning that was validated in prospective [2] or retrospective [3] analysis of clinical trials. However, most clinical trials evaluate one biomarker and one targeted therapeutic at a time. This limits the ability to make high-confidence clinical decisions in a real world scenario with a large number of biomarkers to measure and potential treatments to choose from.

The consequences of strict therapeutic "if..when" protocols in precision medicine can be (I) high selectiveness - where only a small number of patients are assigned to a treatment with robust clinical evidence, (II) high compassionate use - where a majority of patients are left with limited treatment options that, if treated outside of a therapeutic protocol, do not contribute systematically to the development of new clinical evidence and (III) limited predictive ability - where due to a reduced number of new observations, therapeutic protocols can not improve beyond their initial version. For example, in recent precision medicine trials less than 50% of all screened patients were assigned to a treatment [6] and only a third of treated patients showed signs of increased progression-free-survival relative to their prior treatment [7].

Contextual bandits are a class of problems in the field of policy learning in which every action of the agent leads to a direct reward, depending on the state the agent is in. However, in contrast to more complex reinforcement learning problems, the agent does not have to develop a strategy that spans multiple action-state pairs. What differentiates contextual bandit problems from simple classification tasks, is that the agent learns a policy over time and thus has to balance exploration and exploitation to maximize the reward function. A popular analogy is the visit of a casino. The agent is given 10 minutes to play and has a set of multiple slot machines (bandits) to chose from. Every slot machine comes with a different probability to return a reward, however, this information is hidden. The goal of the agent is to maximize the reward. Thus, in the next 10 minutes the agent has to find a balance between the costs of estimating the winning probability for every slot machine (exploration) and focusing on the most rewarding slot machine (exploitation).

Submitted to 32nd Conference on Neural Information Processing Systems (NIPS 2018). Do not distribute.

In this class project we aim to formalize the administration of targeted therapeutics in oncology as a contextual bandit problem [10]. We have curated a public dataset of 1000 cancer cell lines and their response to 8 different targeted therapeutics. For every cell line we have aggregated mutation, CNV and gene expression data. We will use Bayesian neural networks to subsequently chose the best treatment for a randomly selected cell line by observing the genomic data. Based on the treatment decision, the agent will be rewarded with the in-vitro drug response the cancer cell line showed to the agent. Thus, the agent will try to learn the genomic predictors of drug response in the most efficient way possible. In parallel, we have collected a set of therapeutic protocols for each compound that reflect current clinical evidence. During this project, we will try to integrate this prior knowledge into the available state information to facilitate the agent's learning.

The presented problem has potential real-world implications for precision oncology. While in-vitro drug sensitivity screening of cancer cells can test every possible treatment response and link it to genomic predictors, this is not the case in precision oncology. Here, an individual patient is treated with a single drug and thus, only one potential outcome is realized and can be observed. Identifying strategies in which patients can be allocated to treatment options so that the overall patient benefit is maximized is a fundamental goal of precision oncology.

2 Related Work

Recently the number of targeted therapeutics tested in clinical trials has grown rapidly, prompting the introduction of modern master trial designs, such as baskets, umbrellas and platforms, thereby improving the throughput of efficacy testing [8]. Of note, first master protocols such as the I-SPY2 trial for neoadjuvant breast cancer therapy have employed an adaptive design based on bayesian hierarchical models to guide treatment assignment and treatment arm discontinuation [4, 9]. Despite these changes in clinical trial design, most studies still stratify patients by a limited number of biomarkers and strictly adhere to pre-defined treatment arms.

Deep learning based methods have been used for the optimization of contextual bandit problems in the past and reach competitive performance to linear methods in most benchmarks in a Thompson sampling framework. As state information becomes more complex, the expressivity of linear methods shows clear limitations. Here we are going to use the author's public code base to use bayesian neural networks for policy learning. We will compare the algorithms performance against other benchmarks, such as random allocation, Gaussian processes and, tailored to our scenario, current clinical evidence.

3 Dataset

We synthesized a public dataset of in-vitro cancer cell-line drug sensitivity for the purpose of causal inference evaluation [5]. In this dataset a complete matrix of treatment effects for >1000 cell-lines and >50 drugs has been measured and the IC50 values were recorded. Moreover, the dataset contains gene expression data as well as mutation status for most cell-lines.

In a first pre-processing step we focused on a set of 10 drugs that are currently used in clinical practice. We then log transformed the $IC50$ values and normalized them relative to the median $ln(IC50)$ across cell-lines for each drug.

As we are comparing the effectiveness of potential outcomes on a cell-line level, the latter step might seem counter-intuitive. However, the pharmacokinetics and phamacodynamics of therapeutic substances ex-vivo does, in general, not match in-vivo conditions. Thus, it is safer to estimate a treatment's in-vitro effect by comparing the transformed $IC50$ of a drug in a given cell-line relative to the median transformed $IC50$ of the same drug across the whole population of cell-lines.

Next we manually curated therapeutic protocols based on current clinical evidence and trial protocols with selected simplifications: (I) we excluded any protocols involving combination treatments, (II) we excluded any protocols that are based on the presence of oncogenic gene-fusions, (III) we did not include tissue type restrictions into any protocols.

For the start, we propose to reduce the complexity of the cell-line features X into larger subgroups defined by (I) tissue type and (II) mutation status. Depending on the time, we will try to incorporate gene expression data as well.

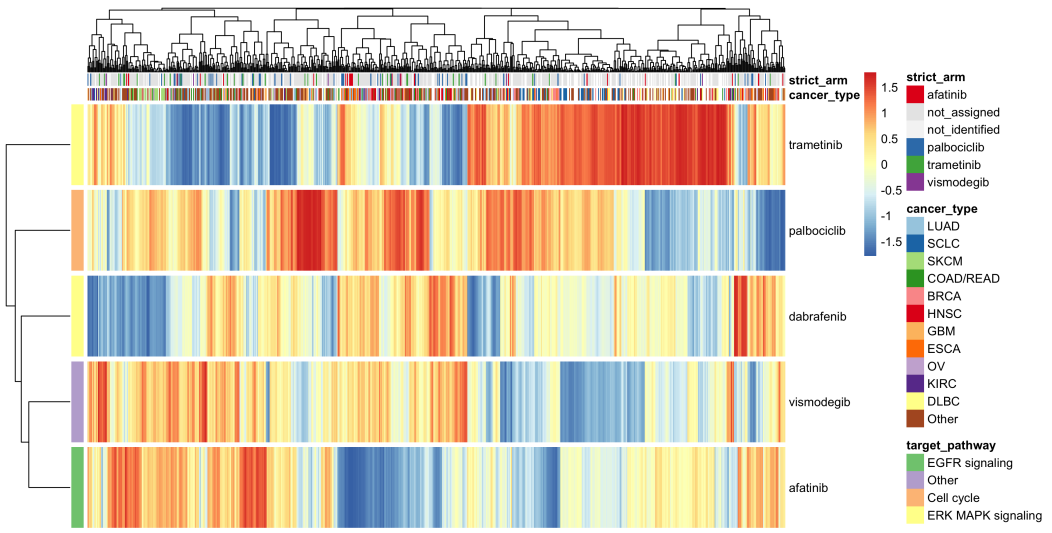

Figure 1: Drug vulnerability data for 1000 cell lines and 5 drugs. Blue color to drug sensitivity. The "strict arm" represents a cell line's treatment allocation based on current clinical evidence if it was a patient in a precision oncology program.

4 Disclosure

Niklas Rindtorff is working with the same curated dataset on a class project in CS282R, a Harvard run course in collaboration with Google Brain. Niklas Rindtorff has manually pre-processed the data and prepared it for use in causal inference of conditional treatment effects. This proposal includes the following new elements of work, which are unrelated to any other projects:

- Formatting and filtering the dataset to be amenable for a contextual bandit learning
- Potential dimensionality reduction of feature data
- Creating a contextual bandit environment
- Adding prior knowledge, in the form of therapeutic protocols to the state information
- Train a Bayesian neural network and other models in a bandit environment
- Evaluate the performance of the final model in the context of current clinical evidence. We will assess the number of cell lines treated with treatments that cause a viability response as well as the average treatment effect accross cell lines that have been treated by both the agent and the current clinical evidence.
- Aggregate and present findings

References

[1] David Blumenthal and Marilyn Tavenner. "A New Initiative on Precision Medicine". In: *The New England journal of medicine* 363.1 (2010), pp. 1–3. ISSN: 15334406. DOI: 10.1056/NEJMp1002530. arXiv: arXiv:1011.1669v3. URL: http://scholar.google.com/scholar?hl=en%7B%5C&%7DbtnG=Search%7B%5C&%7Dq=intitle:New+engla+nd+journal%7B%5C#%7D0.

[2] Jeannie Hou et al. "Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation". In: *New England Journal of Medicine* 364.26 (2011), pp. 2507–2516. ISSN: 0028-4793. DOI: 10.1056/nejmoa1103782.

[3] Jonathan Ledermann et al. "Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: A preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial". In: *The Lancet Oncology* 15.8 (2014), pp. 852–861. ISSN: 14745488. DOI: 10.1016/S1470-2045(14)70228-1. URL: http://dx.doi.org/10.1016/S1470-2045(14)70228-1.

[4] Rajeshwari Sridhara et al. "Current Statistical Challenges in Oncology Clinical Trials in the Era of Targeted Therapy". In: *Statistics in Biopharmaceutical Research* 7.4 (2015), pp. 348–356. ISSN: 19466315. DOI: 10.1080/19466315.2015.1094673.

[5] Francesco Iorio et al. "A Landscape of Pharmacogenomic Interactions in Cancer". In: *Cell* 166.3 (2016), pp. 740–754. ISSN: 10974172. DOI: 10.1016/j.cell.2016.06.017. arXiv: NIHMS150003.

[6] Anthony Letai. *Functional precision cancer medicine-moving beyond pure genomics*. 2017. DOI: 10.1038/nm.4389. URL: http://dx.doi.org/10.1038/nm.4389.

[7] Christophe Massard et al. "High-throughput genomics and clinical outcome in hard-to-treat advanced cancers: Results of the MOSCATO 01 trial". In: *Cancer Discovery* 7.6 (2017), pp. 586–595. ISSN: 21598290. DOI: 10.1158/2159-8290.CD-16-1396.

[8] Janet Woodcock and Lisa M LaVange. "Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both." In: *The New England journal of medicine* 377.1 (2017), pp. 62–70. ISSN: 1533-4406. DOI: 10.1056/NEJMra1510062. URL: http://www.ncbi.nlm.nih.gov/pubmed/28679092.

[9] Shinjo Yada and Chikuma Hamada. "Application of Bayesian hierarchical models for phase I/II clinical trials in oncology". In: *Pharmaceutical Statistics* 16.2 (2017), pp. 114–121. ISSN: 15391612. DOI: 10.1002/pst.1793.

[10] Carlos Riquelme, George Tucker, and Jasper Snoek. "Deep Bayesian Bandits Showdown: An Empirical Comparison of Bayesian Deep Networks for Thompson Sampling". In: (2018). arXiv: 1802.09127. URL: http://arxiv.org/abs/1802.09127.