CS 234, Winter 2023
Stanford University

Default Final Project
Estimation of the Warfarin Dose

Eva Lestant, Amar Venugopal
elestant@stanford.edu, amarvenu@stanford.edu

Abstract

Warfarin is the most widely used oral blood anticoagulant agent worldwide, with **over 30 million prescriptions** in the United States alone in 2004. However, prescribing the appropriate dosage of Warfarin for each patient can be challenging due to **significant individual variability**. Incorrect dosages can result in **severe consequences**, including dangerous bleeding or inadequate prevention of blood clots. While various approaches, such as pharmacogenetic and clinical dosing algorithms, have been developed to determine the initial dosage, they still **rely on a trial-and-error procedure**, which can lead to adverse effects.

To address this challenge, this project aims to apply **multi-armed bandit algorithms** to predict the correct dosage of Warfarin without relying on a trial-and-error procedure. Specifically, the project investigates the performance of linear bandit algorithms, demonstrating that **LinUCB outperforms both the fixed and clinical dosing baselines**. We also investigate alternative model formulations and assess performance.

Data

This project leverages a publicly available patient dataset collected by PharmGKB. Comprising of **5528 patients drawn from studies across 9 countries**, this dataset includes optimal patient-specific warfarin doses, as well as **patient features** such as gender, race, height, weight, medical history, genotypes, and phenotypes.

There are, however, a significant number of entries for which **data is missing**. We impute missing values of VKORC1, a genotype feature, based on the algorithm provided in the appendix to the dataset (I.W.P, 2009). The final set of features that we consider are **age, height, weight, race, enzyme inducer status, amiodarone use**, VKORC1, and CYP2C9 (another genotype feature, for which an imputation algorithm is not provided). We then drop all observations for which any of these features are null while treating unknown values of VKORC1 and CYP2C9 as a separate feature class. The result leaves us with 4386 observations with full data.

In order to handle these categorical genotypic features, we employ **one-hot encoding,** constructing a dummy variable for each class membership. We then drop one such dummy variable from each group and include an intercept in our feature space, in keeping with standard practice for linear models. **The end result is a dataset with 4386 observations and 23 features.**

The label, which we try to predict, is given by **the correct therapeutic dose** of warfarin, which we discretize into **3 bins** corresponding to "*low*", "*medium*", and "*high*".

Models

Benchmark Models

We utilize two benchmark models against which we evaluate bandit performance. The first is the **fixed dose** baseline, which simply assigns every patient a medium dose. The second is the **clinical dosing** algorithm, a simple equation that calculates the square root of weekly dose as a linear function of patient covariates:

$$\begin{aligned}&\sqrt{\text{weekly dose}}\\&= 4.0376 - 0.2546 * \text{Age} + 0.0118 * \text{Height (cm)} + 0.0134 * \text{Weight (kg)}\\&- 0.6752 * \text{Asian} + 0.4060 * \text{Black} + 0.0443 * \text{Missing or mixed race} + 1.2799\\&* \text{Enzyme inducer status} - 0.5695 * \text{Amiodarone status}\end{aligned}$$

LinUCB with Disjoint Linear Models

Our primary bandit algorithm is **LinUCB with disjoint linear models**, based on **Li et al. (2010).** In this model, the optimal patient dose is selected based on **a linear function of the patient's characteristics** plus an **optimism term** to ensure sufficient exploration. In the below equations, **D** refers to the matrix of past covariates for patients assigned to arm *a*.

$$a_t = \arg\max_{a \in \mathcal{A}_t} \left(x_{t,a}^\top \hat{\boldsymbol{\theta}}_a + \alpha \sqrt{x_{t,a}^\top \boldsymbol{A_a^{-1}} x_{t,a}} \right)$$

Where $\boldsymbol{A_a} = \boldsymbol{D_a^\top D_a} + I_d$

Thompson Sampling

Our second bandit algorithm is **Thompson Sampling**, applied to **contextual bandits** with linear payoffs, based on **Agrawal & Goyal (2013)**. This algorithm works **by posterior sampling**, where we sample the linear parameter vector from a continuously **updating normal distribution**, We deviate from the prescribed implementation by giving each arm a separate set of parameters, thereby making it a disjoint setting similar to LinUCB.

$$B_a(t) = I_d + \sum_{\tau=1}^{t-1} b(\tau) b(\tau)^T \, 1\{a_\tau = a\}$$

$$\hat{\mu}_a(t) = B_a(t)^{-1} \left(\sum_{\tau=1}^{t-1} b(\tau) r(\tau) 1\{a_\tau = a\} \right)$$

$$\tilde{\mu}_a(t) \sim \mathcal{N}(\hat{\mu}_a(t), v^2 B_a(t)^{-1})$$

$$a_t = arg\max_a b^T(t) \tilde{\mu}_a(t)$$

Supervised Linear Bandit

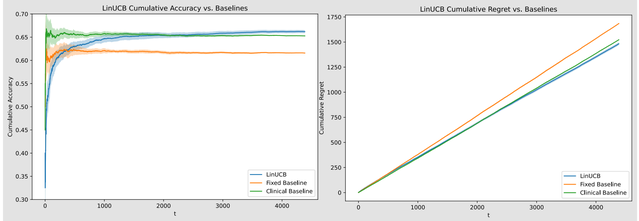
Our final algorithm is a **supervised linear bandit**, which trains a linear model based on all prior observed **true** Warfarin doses, rather than the standard binary reward. By utilizing this extra information, this model should achieve better performance than all others tested and will set a kind of upper bound on the performance of linear bandit models.

$$\hat{\theta}_t = (X_t^T X_t)^{-1} X_t^T Y_t$$

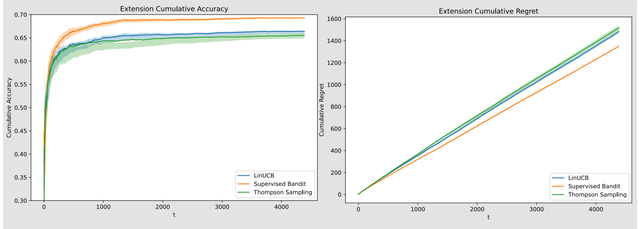
$$a_{t+1} = \arg\max_{a \in \{0,1,2\}} ||a - x_{t+1}^T \hat{\theta}_t||$$

Results

We find that LinUCB significantly **outperforms** the fixed dose baseline, both in terms of **both cumulative accuracy and cumulative expected regret.** The resulting performance plots are as follows, with the shaded regions corresponding to 95% confidence intervals over the 20 different random patient orderings tested:



We also go on to find that our extension models, namely **supervised linear regression "bandit"** and **Thompson Sampling**, achieve good performance:



We can see that Thompson Sampling appears to achieve similar performance to LinUCB (within error). As expected, the supervised linear "bandit" achieves significantly higher accuracy and lower regret compared to both models. This is **unsurprising**, as it is given **access to the true treatment** applied to past patients, rather than the reward, which amounts to just a binary indicator for correct/incorrect treatment.

Discussion & Conclusion

Our results show clear performance improvements when comparing bandit methods to both the fixed dose baseline and, in the case of LinUCB, to the clinical dosing algorithm as well. This clearly demonstrates **the power of bandit methods** for Warfarin dose assignment and suggests that such methods should be **further explored** to achieve better patient outcomes.

However, there is one key caveat to these results: while **LinUCB outperforms the clinical dosing** algorithm by the end of the "online" training period, **it takes nearly 2000 patients** before its performance truly matches that of the clinical dosing algorithm. As a result, if this were a truly online setting, **several of those patients may have received a lower quality of care** than if the clinical dosing algorithm had been applied. This raises **ethical considerations** regarding potential harm to patients used in training bandit methods, particularly **for healthcare applications**. This suggests that **some degree of offline learning**, such as what we have performed in this project, is necessary prior to the rollout of a bandit algorithm in such contexts.

Future

The key direction of future research in this project is to investigate **further implementations of linear bandit models** to attempt to close the performance gap between our current optimal implementation, **LinUCB, and the supervised linear bandit**. While the latter's performance presents an upper bound on the potential performance of linear bandit models and likely **cannot be replicated** by a model that observes only binary rewards, the existing gap suggests that **other model formulations may be able to provide better performance**.

Possible candidates that we have not tested in this project include **robust algorithms or regularized linear models**.

References

I. W. P. Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. *New England Journal of Medicine*, 360(8):753–764, 2009.

L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.

S. Agrawal and N. Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pages 127–135, 2013.