

Parameter Sweep Review

Tianze (Steven) Shou

2022-10-10

Introduction

Starting from November, a revisit on the parameter sweep algorithm will be initiated. Therefore, this report will serve as a review on the concepts presented and work done for parameter analysis in Summer 2022. Some discovered issues for the algorithm will be discussed. In the meantime, this report will also propose some viable ways to improve or restructure the parameter sweep algorithm.

Current Parameter Sweep Structure

Data Sources

The current parameter sweep algorithm takes data from two sources: real-time Pozyx teacher position log and event coding by human observer in the classroom (AKA Shamyia).

From these two datasets, we are interested in investigating the following question: how can we configure the parameters for **stops** such that the Pozyx position log can best predict teacher's stopping/visiting individual student behavior?

However, we are working with some innate flaws that these datasets carry. Since the Pozyx system carries some amount of sensing inaccuracy, the position log is produced with a *confidence score*. Investigation on this statistic is accessed by Wenhan, whose work could be potentially involved in improving parameter sweep. Shamyia's input will be needed for further discussion.

Now we suppose one of the true teacher stops takes place at time stamp t_0 and ends at t_1 , and the observation log denotes this event at time t^* . This coding has two issues: (1) $t^* = t_0 + \epsilon$ where $\epsilon > 0$ is the amount of human observer lag, and we do not know its distribution; (2) our estimate for t_1 cannot be obtained directly from the observation log. Having an estimate on the ending time stamp of a stop is important since it is necessary for assessing the algorithm's accuracy. Therefore, if we are going to use the human observer log as target variable, these two issues will have to be addressed.

Algorithm Structure

For the current version of the algorithm, we defined two sets of parameters. **Hyperparameters:** **reward**, **penalty**, **timeframe**, etc.; **model parameters:** **radius**, **duration**, and **range**. We would first input the feasible regions and step sizes for each of the hyperparameters and parameters, followed by a grid search conducted by the algorithm. This computing method is fundamentally costly and slow. Although it does not affect the eventual outcome of the best parameter set, it still hinders the development pace. In the following section, a proposal methods for improvement will be proposed.

Proposed Changes for Future Iterations

For issue (2) in the observation log, we might be able to extract t_1 for each stop from the starting time stamp of the next event. Since there is only one teacher and he/she can only be involved in one event at a time.

The starting time stamp of the next event will be a rather accurate approximation of the ending time of the current stop.

To improve the running time for the algorithm, we are going to keep the grid search structure for the hyperparameters, but the search for best model parameters can be achieved through techniques like gradient descent, since the optimization problem is convex according to my observation with previous results.