

Table 1: Table shows the result of the linear mixed model. It uses viability as dependent variable and the hyperparameters as independent numerical variables. The model is adjusted for general differences in individual hyperparameter settings.

	Coef.	Std.Err.	p-value
Intercept	2.523	0.069	0.000
erate	0.009	0.011	0.443
erate:dmrate	-0.158	0.012	0.000
erate:imrate	-0.009	0.012	0.463
erate:cmrate	-0.072	0.011	0.000
erate:tmrate	-0.196	0.011	0.000
Hyperparameter Set Var	0.029	0.114	

0.1 Determine the Evolutionary Algorithm Mutation-Rates

0.1.1 Experimental Setup

Given the set of configurations chosen in ??, we will explore optimal parameter setting. Therefore, we run each configuration with a multitude with mutation rates.

For each configuration set we uniformly sample the mutation rates between 0 and 1. We use the resulting viability and feasibility as dependent variables and each hyperparameter edit-type in interaction with edit-rate as independent variable. The remaining procedure follows the process described in ??.

0.1.2 Results

[Correlation matrix makes no sense. It should be only a bar plot with feasibility and viability with all the hyperparams.]

Figure 1 suggests that all hyperparameters have a negative correlation with feasibility and viability. Even, if we consider the interaction with edit-rate. The edit-rate itself appears to be beneficial.

If we inspect the model results in terms of viability Table 1, we see a similar trend. Furthermore, it appears, that the specific hyperparameter set affects the results positively. This suggests, that the edit-rate might be the most important hyperparameter. Secondly, we can assume that a carefully chosen balance in hyperparameters to achieve an increase in viability.

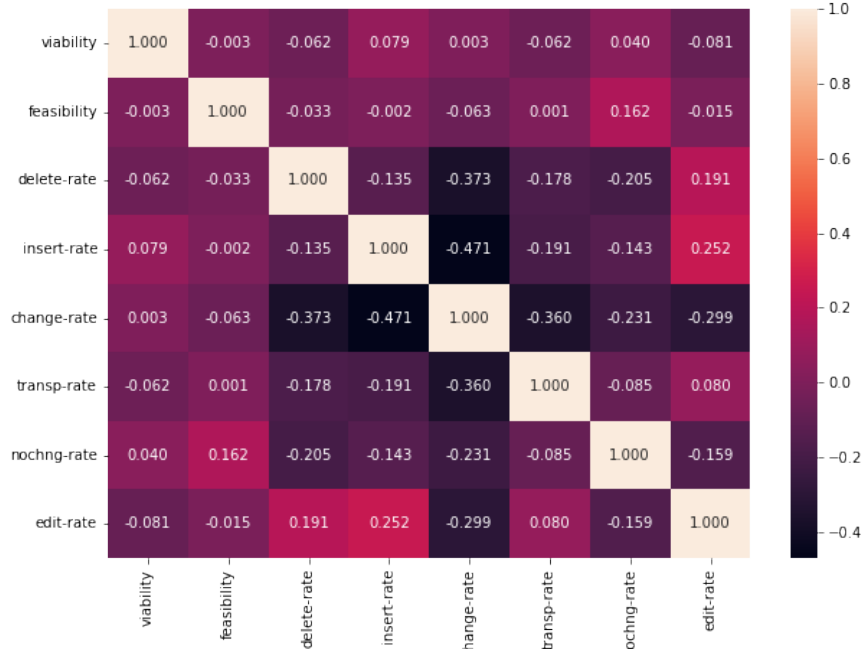


Figure 1: Shows a correlation matrix of each hyperparameter, the feasibility measure and the viability measure.

Table 2: Table shows the result of the linear mixed model. It uses feasibility as dependent variable and the hyperparameters as independent numerical variables. The model is adjusted for general differences in individual hyperparameter settings.

	Coef.	Std.Err.	p-value
Intercept	0.022	0.011	0.041
erate	0.099	0.004	0.000
erate:dmrate	-0.105	0.004	0.000
erate:imrate	-0.105	0.004	0.000
erate:cmrate	-0.107	0.004	0.000
erate:tmrate	-0.102	0.004	0.000
Hyperparameter Set Var	0.001	0.009	

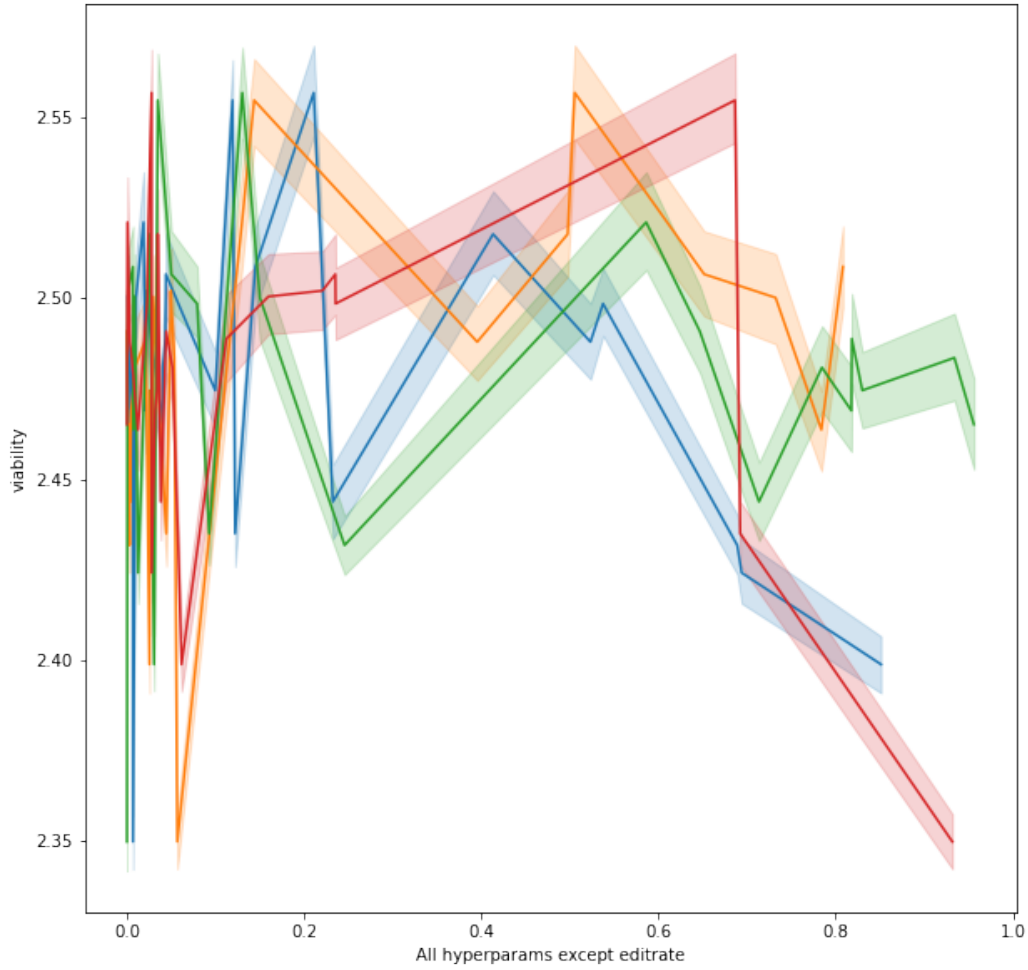


Figure 2: Shows all 4 edit-types and their effect on viability.

Looking at feasibility in Table 2, we see the same behavior. However, the models defined in this section only explain linear behaviours. Hence, Figure 2, show each individual effect in relation to viability.

Figure 2 indeed shows that the edit-types highly fluctuate in their effect on viability. These results suggest that the linear models do not accurately reflect the effect of each of the edit-types. [FOR XIXI: I know this plot looks horrible. I am sorry, I couldn't polish it on time.]

0.1.3 Discussion

[FOR XIXI: This section may not be correct anymore as I couldn't manage to rerun the experiment on time. These remarks are based

on older results.] The result is reasonable, as most counterfactuals that differ from the initial factual have never been seen in the dataset. Therefore, they are deemed impossible by the feasibility measure. With this in mind, we opt to choose the set of hyperparameters that increase the chance of finding feasible counterfactuals by averaging the hyperparameters of the non-zero feasibility subset. This selection most likely decreases viability, but this is preferred over impossible counterfactuals.

Therefore, we choose **[a set of specific hyperparameter values]** for any subsequent experiment.