**Rule-Based Genetic Algorithm toward Explainable AI**

Introduction

With the increased usages of AI, AI accountability is receiving keen attention from the society. Deep learning has enhanced accuracy of predictions and helped many domains utilize its wealth of the technological advantages, while non-linear complex model lacks interpretable self-explanations. In this research we aim for building an explainable AI model using the rule-based genetic algorithm.

Methodology

**Genes**

Each gene is represented by a set of rule-bundles and each rule-bundle consists of rules with four respective intervals for attributes of Iris – Sepal Length/Width, Petal Length/Width – and three different probabilities for a sample Iris to be respective Iris type – Setosa, Versicol, Virginica. If the sample Iris satisfies the rule, i.e., all of its attributes fall within respective intervals, then the predictor will return three respective probabilities for three Iris types.

Diagram

Description automatically generated with medium confidence

Fig: Rule

Diagram

Description automatically generated

Fig: Rule-Bundle

**Intervals**

Intervals are generated from the training dataset so that each interval contains the same number of data points. In case the number of the training dataset is not divisible by an intended number of intervals, such as 10 intervals for 101 input data, then one interval will be randomly selected and assigned the extra data point, i.e., one interval receives 11 input data falls, whilst others receive 10.

The lowest interval and the highest intervals contain infinite boundaries for lower end and upper end of the intervals respectively. It will most likely not beneficial for Iris data since none of attributes may be negative values and/or we will not see astronomically huge value for the upper end. However, it is designed this way for the sake of future when we utilize this model for other domains that deal with negative/huge values.

**Evolutionary Process**

The evolutionary process takes place for X generations. It begins with N number of first-generation rule-bundles and each of rule-bundles contains M number of rules. The algorithm first evaluates all rule-bundle in the first generation by comparing aggregated losses of rules contained in each rule-bundle. A better performing rule-bundle replaces the previous best bundle.

The subsequent evolutionary process takes place in iteration. Amongst all the rules of a current generation, *n\_parents* number of rule-bundle2 (20 times in the below experimentation) are randomly selected as a parent’s pool, two out of the pool are chosen to cross-over and mutate to generate a child. This generative process occurs *n\_children* number of times (10 times in the below experimentation), and the newly created children replaces equal number of randomly selected parents (In our experimentation, 10 newly created children replace 10/20 selected parents).

**Seed rule-bundle /Rule generation**

The first-generation rule-bundles receive M number of rules generated by the algorithm that randomly picks one of the intervals and generates the same number of weights as class attributes. In Iris case, since there are three types of class, each rule receives 3 weights.

**Crossover**

Crossover slices the two selected parents at anywhere except their head or tail, to avoid generating an identical child rule-bundle to either of the parent rule.

**Mutation**

Some part/all of a rule in each rule-bundle may be mutated at random probability.

**Prediction**

Prediction occurs in two layers; one at each rule level and the other at bundle level. Each rule classifies an Iris into one of three Iris types, unless there is a tie probability, in which case no classification given. Each rule in a bundle votes and the bundle returns the majority as the predicted Iris type.

**Evaluation**

Each bundle is evaluated by the number of accurate predictions divided by the total number of samples/test data. In each generation, the bundle with the highest accuracy score is chosen and compared with the best bundle so far.

Experimentation

*\*Moved this section to other attachment since we ran multiple experimentations this week*