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

Introduction

As AI usage grows, the demand for AI accountability is becoming a significant societal concern. Despite deep learning's success in improving prediction accuracy across various fields, its complex, non-linear models often lack clear interpretability. This research is dedicated to developing an explainable AI model through the use of a rule-based genetic algorithm.

Methodology

**Genes**

Each gene is represented by a set of rule bundles, with each rule bundle comprising rules that define four attribute intervals for the Iris species—Sepal Length, Sepal Width, Petal Length, and Petal Width. Additionally, each rule is associated with three probabilities corresponding to the likelihood of a sample being one of the Iris types—Setosa, Versicolor, or Virginica. If a sample Iris's attributes fall within the specified intervals of a rule, the predictor will output the three corresponding probabilities for the Iris types.

Diagram

Description automatically generated with medium confidence

Fig: Rule

Diagram

Description automatically generated

Fig: Rule-Bundle

**Intervals**

Intervals are derived from the training dataset, with each interval encompassing an equal number of data points. In instances where the number of data points in the training dataset is not evenly divisible by the desired number of intervals, such as attempting to create 10 intervals from 101 data points, one interval will be randomly chosen to accommodate the additional data point. Consequently, this interval will contain 11 data points, while the others will each contain 10.

The lowest and highest intervals are designed with open-ended boundaries at their respective ends. This means the lowest interval has no lower limit, and the highest interval has no upper limit. Such design is unlikely to be advantageous for Iris data, as the attribute values are neither negative nor extremely high. Nevertheless, this approach is intentionally adopted to ensure the model's applicability to other domains in the future, which may involve negative or extremely large values.

**Evolutionary Process**

The evolutionary process unfolds over X generations, starting with an initial population of N first-generation rule bundles, each containing M rules. The algorithm begins by evaluating each rule bundle within the first generation, comparing their aggregated losses. The most effective rule bundle is retained as the best performing one for that generation.

As the process iterates, n\_parents rule bundles are randomly selected from the current generation to form a parental pool, from which two are chosen for crossover and mutation, resulting in a child. This process is repeated n\_children times (demonstrated as 10 times in our experiment), and the resulting offspring replace an equal number of randomly chosen parents. In our experiment, for example, 10 new children would replace 10 out of the 20 parents selected.

**Seed rule-bundle /Rule generation**

The initial generation of rule-bundles is created with M rules each, generated by an algorithm that selects one of the intervals and assigns a corresponding number of weights based on class attributes. In the case of the Iris dataset, which has three classes, each rule is given three weights.

**Crossover**

During crossover, two parent rule-bundles are cut at a point that is neither the beginning nor the end, ensuring the resulting child rule-bundle is not identical to either parent.

**Mutation**

A rule within a rule-bundle may undergo mutation with a certain probability, which could affect some or all of its parts.

**Prediction**

Prediction operates at two levels: the rule level and the bundle level. Each rule attempts to classify an Iris into one of three types. If there is a tie in probability, the rule does not make a classification. Within a bundle, each rule's vote is tallied, and the majority vote determines the predicted Iris type.

**Evaluation**

Rule-bundles are evaluated based on the proportion of accurate predictions out of the total number of samples or test data. In each generation, the rule-bundle with the highest accuracy is identified and compared with the best performing bundle to date.