Coevolutionary Algorithm for Building Robust Decision Trees under Minimax Regret

Adam Żychowski [1] Andrew Perrault [2] Jacek Mańdziuk [1, 3, 4]

[1]Faculty of Mathematics and Information Science, Warsaw University of Technology [2]Department of Computer Science and Engineering, The Ohio State University [3]Faculty of Computer Science, AGH University of Krakow [4]Center of Excellence in Artificial Intelligence, AGH University of Krakow

Scan for arXiv paper

Abstract

Objective: Develop **robust** machine learning models, particularly focusing on **decision trees**.

Algorithm: *CoEvoRDT* - **coevolutionary algorithm** designed for **creating robust decision trees** capable of handling noisy high-dimensional data in adversarial contexts. CoEvoRDT alternately evolves **competing populations of decision trees and perturbed features**, facilitating the construction of decision trees with desired properties.

- **Game Theory Inspiration**: utilizes mixed Nash equilibrium to enhance convergence during the coevolution process.
- **Potential to improve results of other state-of-the-art methods** by incorporating their outcomes (decision trees they produce) into the initial population and optimizing them through coevolution.
- **Tested on 20 popular datasets**, demonstrating superior performance compared to 4 state-of-the-art algorithms.
- **Easily adaptable to various target metrics**: allowing the use of tailored robustness criteria such as minimax regret.
- **Adversarial Accuracy**: outperformed all competing methods on 13 datasets with adversarial accuracy metrics.
- **Minimax Regret**: achieved superior performance on all 20 considered datasets with minimax regret as the evaluation metric.

Problem definition

Let $X \subset \mathbb{R}^d$ be a d-dimensional instance space (inputs) and Y be the set of possible classes (outputs).

A classical classification task is to find a function (model) $h : X \to Y$, $h(x_i) = y_i$, where y_i is true class of x_i. Classification performance of model h can be measured by accuracy:

$$\mathrm{acc}(h) = \frac{1}{|X|} \sum_{x_i \in X} I[h(x_i) = y_i],$$

where $I[h(x_i) = y_i]$ returns 1 if h predicts the true class of x_i, and 0, otherwise.

Let $\mathcal{N}_\varepsilon(x) = \{z : ||z - x||_\infty \leq \varepsilon\}$ be a ball with center x and radius ε under the L_∞ metric. The **adversarial accuracy** of a model h is accuracy on the perturbation in the perturbation set that produces the lowest accuracy. It is formally defined as

$$\mathrm{acc}_{\mathrm{adv}}(h, \epsilon) = \frac{1}{|X|} \sum_{x_i \in X} \min_{z_i \in \mathcal{N}_\varepsilon(x_i)} I[h(z_i) = y_i].$$

The **max regret** of a model h is the maximum *regret* among all possible perturbations $z \in \mathcal{N}_\varepsilon$. Regret is the difference between the best accuracy possible on a particular perturbation and the accuracy h achieves:

$$\mathrm{regret}(h, \{z_i\}) = \max_{h'} \mathrm{acc}(h', \{z_i\}) - \mathrm{acc}(h, \{z_i\}),$$

where $\mathrm{acc}(h, \{z_i\})$ is the accuracy achieved by h when $\{x_i\}$ is replaced with $\{z_i\}$. Max regret be expressed as:

$$\mathrm{mr}(h) = \max_{z_i \in \mathcal{N}_\varepsilon(x_i)} \mathrm{regret}(h, \{z_i\}).$$

The problem is **finding a decision trained on X that for a given ε optimizes a given robustness metric** (maximizes for adversarial accuracy or minimizes for max regret).

Example

T1

	CS	*I*	*D*
A1	50	10	0
A2	60	20	1
A3	60	30	1

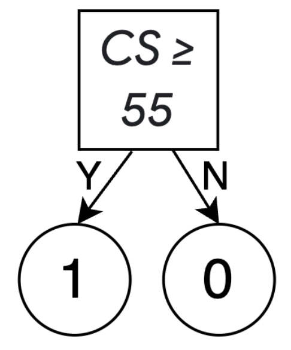
T2

	CS	*I*	*D*
A1	55 (+5)	10	0
A2	55 (-5)	20	1
A3	60 (+0)	30	1

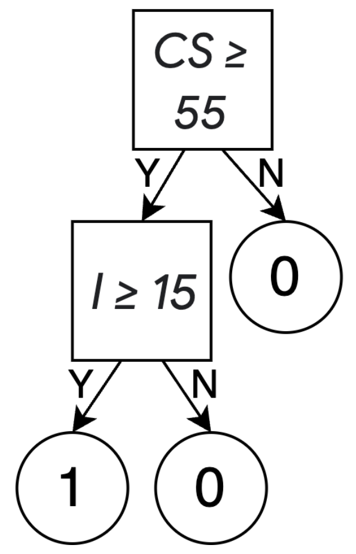
T3

	CS	*I*	*D*
A1	55 (+5)	20 (+10)	0
A2	55 (-5)	20 (+0)	1
A3	55 (-5)	20 (-10)	1

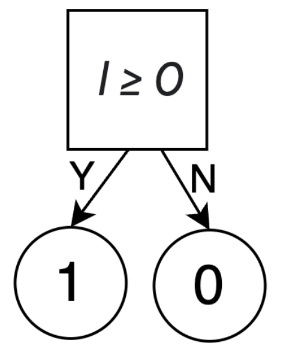
DT1



DT2



DT3

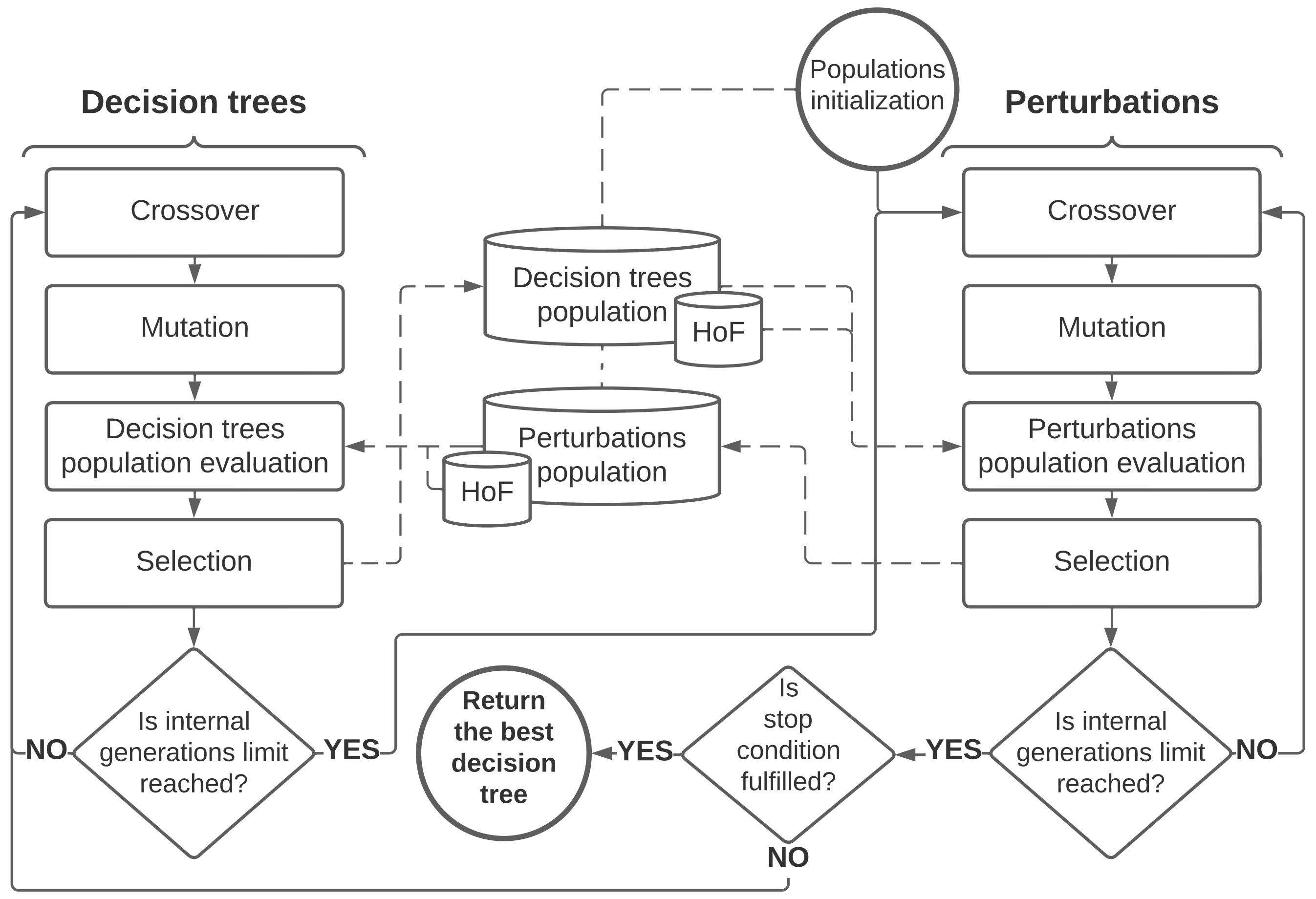


CoEvolutionary method for Robust Decision Trees (CoEvoRDT)

Overview

CoEvoRDT maintains **two populations**: one contains **encoded decision trees**, and the other contains **input data perturbations**. Both populations are initialized with random elements and then developed alternately. First, the decision tree population is modified by evolutionary operators (crossover, mutation, and selection) through given number of generations. Then, the perturbation population is evolved through the same number of generations. The above loop is repeated until the stop condition is satisfied.

Algorithm workflow


Populations initialization

Decision trees

Perturbations

Crossover

Mutation

Decision trees population evaluation

Selection

Decision trees population

HoF

Perturbations population

HoF

Crossover

Mutation

Perturbations population evaluation

Selection

Is internal generations limit reached?

NO

YES

Return the best decision tree

YES

Is stop condition fulfilled?

YES

NO

Is internal generations limit reached?

NO

Decision Tree Population

- Each **decision tree is encoded as a list of nodes**, where each node is represented by a 7-tuple $\{t, c, P, L, R, o, v, a\}$: node number ($t$), class label ($c$), parent node pointer ($P$), left and right children pointers (L and R), operator indication (o), value to be tested (v), and attribute (a).
- **Initial population:** random decision trees with depth between 2 and 10.
- **Crossover:** occurs with a probability, randomly pairing individuals and exchanging entire subtrees between selected nodes to generate offspring.
- **Mutation:** applied with a probability, introducing random changes through actions like subtree replacement, node information change, or subtree pruning.
- **Evaluation procedure:** performed against the perturbation population. Metric optimization is computed against all perturbations from the adversarial population.

Perturbation population

- Each individual represents a **perturbed input set** X, with perturbations constrained within ϵ.
- **Initial population:** Random perturbations generated uniformly, meeting ϵ criteria.
- **Crossover:** Selects a random subset of individuals, pairs them randomly, and mixes perturbed input instances from both parents to generate offspring.
- **Mutation:** Independently perturbing each input instance's encoded values.
- **Evaluation:** Balance perturbation efficiency against all decision trees and avoiding oscillation. $N_{top} = 20$ highest-fitness decision trees are used for perturbation evaluation.

Hall of Fame and Nash Equlibrium

- **Role:** Mechanism to retain and store best-performing individuals encountered during evolution.
- **Common approach critique:** Traditional approach adds one highest-fitness individual per generation, potentially suboptimal for diversity.
- **CoEvoRDT approach:** Utilizes a game-theoretic approach treating decision trees and perturbations as strategies in a non-cooperative zero-sum game.
- **Mixed Nash Equilibrium:** Calculates mixed Nash equilibrium, resulting in mixed strategies for both decision trees and perturbations.
- **Evaluation enhancement:** Fitness function calculated against a merged set of Hall of Fame and population individuals.
- **Size limitation:** Hall of Fame size is limited, with the lowest-fitness element determining the limit.

Main results

dataset	CART	Meta Silvae	RIGDT-h	GROOT	FPRDT	CoEvoRDT	**CoEvoRDT+FPRDT**
ionos	.094±.000	.075±.007	.071±.006	.061±.005	.061±.006	**.052±.004**	.052±.005
breast	.103±.000	.056±.006	.069±.006	.059±.005	.057±.005	**.049±.004**	.049±.005
diabetes	.202±.000	.126±.008	.132±.009	.124±.009	.117±.007	**.096±.006**	.094±.007
bank	.186±.000	.102±.007	.108±.008	.090±.006	.089±.007	**.076±.006**	.076±.006
Japan3v4	.107±.000	.090±.006	.083±.006	.067±.006	.066±.004	**.062±.006**	.061±.006
spam	.097±.000	.079±.006	.083±.006	.074±.006	.074±.006	**.070±.005**	.069±.005
GesDvP	.152±.000	.129±.008	.133±.010	.129±.008	.131±.009	**.114±.007**	.114±.007
har1v2	.105±.000	.074±.006	.084±.007	.068±.006	.068±.006	**.064±.005**	.064±.005
wine	.140±.000	.125±.008	.127±.009	.111±.009	.109±.008	**.090±.006**	.090±.007
collision-det	.142±.000	.099±.007	.093±.007	.088±.006	.091±.007	**.061±.006**	.059±.006
mnist:1v5	.249±.000	.078±.007	.076±.006	.071±.006	.067±.005	**.055±.006**	.055±.005
mnist:2v6	.268±.000	.083±.007	.087±.006	.072±.005	.069±.005	**.055±.004**	.054±.004
mnist	.395±.000	.143±.009	.139±.009	.125±.007	.124±.009	**.113±.008**	.112±.008
f-mnist2v5	.273±.000	.254±.015	.249±.015	.223±.013	.238±.014	**.196±.011**	.196±.011
f-mnist3v4	.290±.000	.259±.014	.254±.015	.246±.014	.232±.013	**.202±.011**	.199±.011
f-mnist7v9	.283±.000	.255±.014	.251±.015	.237±.014	.240±.014	**.208±.013**	.207±.012
f-mnist	.427±.000	.345±.020	.337±.018	.292±.017	.286±.016	**.238±.014**	.237±.015
cifar10:0v5	.419±.000	.351±.019	.379±.021	.347±.019	.314±.018	**.241±.015**	.236±.013
cifar10:0v6	.403±.000	.362±.021	.368±.020	.342±.018	.341±.019	**.289±.016**	.289±.016
cifar10:4v8	.408±.000	.357±.019	.360±.021	.339±.018	.331±.019	**.283±.016**	.281±.017

Table 1. **Max regrets** (mean ± std error). CoEvoRDT+FPRDT obtained the best results for all datasets. The best results are **bolded**. Gray background indicates that a given method is statistically significantly better than all other methods.

dataset	CART	Meta Silvae	RIGDT-h	GROOT	FPRDT	CoEvoRDT	**CoEvoRDT+FPRDT**
ionos	.310±.000	.695±.039	.701±.045	.783±.047	**.795±.047**	.791±.044	.795±.049
breast	.250±.000	.797±.047	.838±.052	.874±.047	.876±.055	**.885±.054**	.889±.056
diabetes	.542±.000	.554±.035	.569±.033	.623±.043	**.648±.039**	.617±.038	.648±.037
bank	.633±.000	.510±.031	.468±.033	.541±.036	**.658±.040**	.657±.043	.663±.037
Japan3v4	.576±.000	.566±.035	.564±.037	.584±.035	**.667±.039**	.665±.037	.668±.037
spam	.302±.000	.637±.036	.467±.028	.723±.045	.746±.049	**.751±.049**	.753±.045
GesDvP	.478±.000	.637±.039	.548±.033	.716±.045	.735±.040	**.740±.046**	.741±.044
har1v2	.232±.000	.706±.045	.707±.047	.806±.048	.804±.049	**.818±.054**	.820±.052
wine	.620±.000	.637±.039	.474±.027	.637±.036	.674±.037	**.688±.046**	.692±.047
collision-det	.743±.000	.772±.047	.764±.044	.784±.052	.792±.051	**.798±.053**	.803±.049
mnist:1v5	.921±.000	.952±.056	.957±.054	.954±.056	**.966±.058**	.964±.059	.969±.061
mnist:2v6	.862±.000	.906±.054	.919±.050	.917±.052	**.922±.049**	.917±.053	.922±.051
mnist	.673±.000	.702±.041	.704±.042	.743±.048	.742±.049	**.745±.043**	.754±.046
f-mnist2v5	.675±.000	.951±.053	.945±.060	.971±.057	.978±.055	**.982±.055**	.982±.059
f-mnist3v4	.632±.000	.808±.049	.793±.044	.819±.048	.865±.050	**.869±.056**	.870±.054
f-mnist7v9	.642±.000	.824±.045	.81±.052	.829±.052	**.876±.050**	.868±.054	.880±.047
f-mnist	.464±.000	.492±.033	.525±.033	.536±.035	.531±.033	**.544±.036**	.546±.040
cifar10:0v5	.296±.000	.502±.033	.347±.026	.485±.036	.678±.046	**.685±.039**	.693±.039
cifar10:0v6	.587±.000	.540±.038	.477±.029	.556±.037	.688±.040	**.692±.046**	.697±.043
cifar10:4v8	.256±.000	.514±.032	.488±.033	.473±.032	.661±.042	**.663±.045**	.664±.037

Table 2. **Adversarial accuracies** (mean ± std error). CoEvoRDT+FPRDT obtained the best results for all datasets. Box denotes that CoEvoRDT+FPRDT is statistically significantly better than all other methods. The best results are **bolded**. Gray background indicates that a given method is statistically significantly better than all other methods (except CoEvoRDT+FPRDT).

	minimax regret					adversarial accuracy					computation time [s]				
HoF size	Nash mixed tree	Top K as mixed tree	Nash single trees	Top K	Best	Nash mixed tree	Top K as mixed tree	Nash single trees	Top K	Best	Nash mixed tree	Top K as mixed tree	Nash single trees	Top K	Best
0	.261	.261	.261	.261	.261	.533	.533	.533	.533	.533	47	47	47	47	47
10	.242	.248	.247	.251	.259	.535	.535	.535	.534	.533	50	50	50	50	50
20	.240	.246	.245	.249	.256	.536	.536	.536	.536	.534	55	54	55	56	51
50	.241	.244	.245	.249	.254	.536	.536	.536	.536	.534	61	58	59	62	54
100	.239	.243	.243	.247	.253	.538	.538	.537	.537	.535	68	63	66	65	56
200	.238	.242	.242	.244	.250	.543	.539	.540	.539	.535	77	70	76	77	59
500	.237	.241	.241	.243	.248	.545	.540	.540	.540	.536	86	79	91	90	60
∞	.237	.239	.240	.240	.248	.545	.540	.541	.540	.536	86	77	85	85	61

Table 3. Results with respect of HoF size for *fashion-mnist* dataset. ∞ means that there was no limit on HoF size.

Conclusions

- **Novel coevolutionary algorithm for robust decision tree construction.**
- **Adaptable to various target metrics**, suitable for diverse applications, including scenarios combining robustness with other objectives.
- Introduces a game-theoretic approach for constructing the **Hall of Fame using Mixed Nash Equilibrium**, enhancing robustness and convergence speed.
- **Can integrate results from other strong methods** into the initial population for performance improvement.
- Tested on 20 benchmark datasets, **outperforming competitors in minimax regret** and **achieving on-par performance in adversarial accuracy metrics**.
- **Future Work:** Investigating CoEvoRDT as a multi-population algorithm, exploring the potential of the island model for faster convergence and enhanced performance.

The 38th Annual AAAI Conference on Artificial Intelligence, February 2024, Vancouver

adam.zychowski@pw.edu.pl perrault.17@osu.edu jacek.mandziuk@pw.edu.pl