Cultivating Archipelago of Forests: Evolving Robust Decision Trees through Island Coevolution

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

[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

Overview

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

Algorithm: *ICoEvoRDF* - **island-based coevolutionary algorithm** for constructing robust decision tree ensembles.

Problem definition

The **adversarial accuracy** of a model h is accuracy on the perturbation in the perturbation set that produces the lowest accuracy.

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

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 model trained on X that for a given ε optimizes a given robustness metric**.

Example

T1

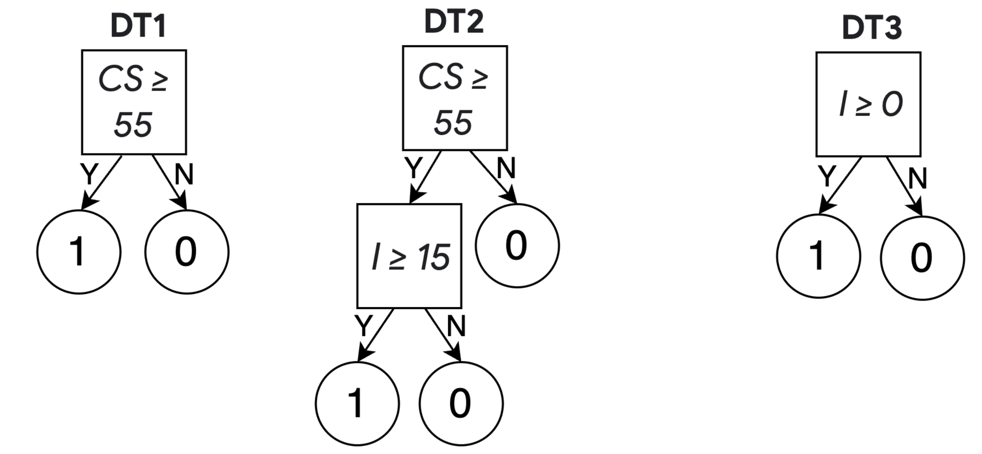
	CS	*I*	*D*
A1	50	10	0
A2	60	20	1
A3	60	30	1

T2

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

T3

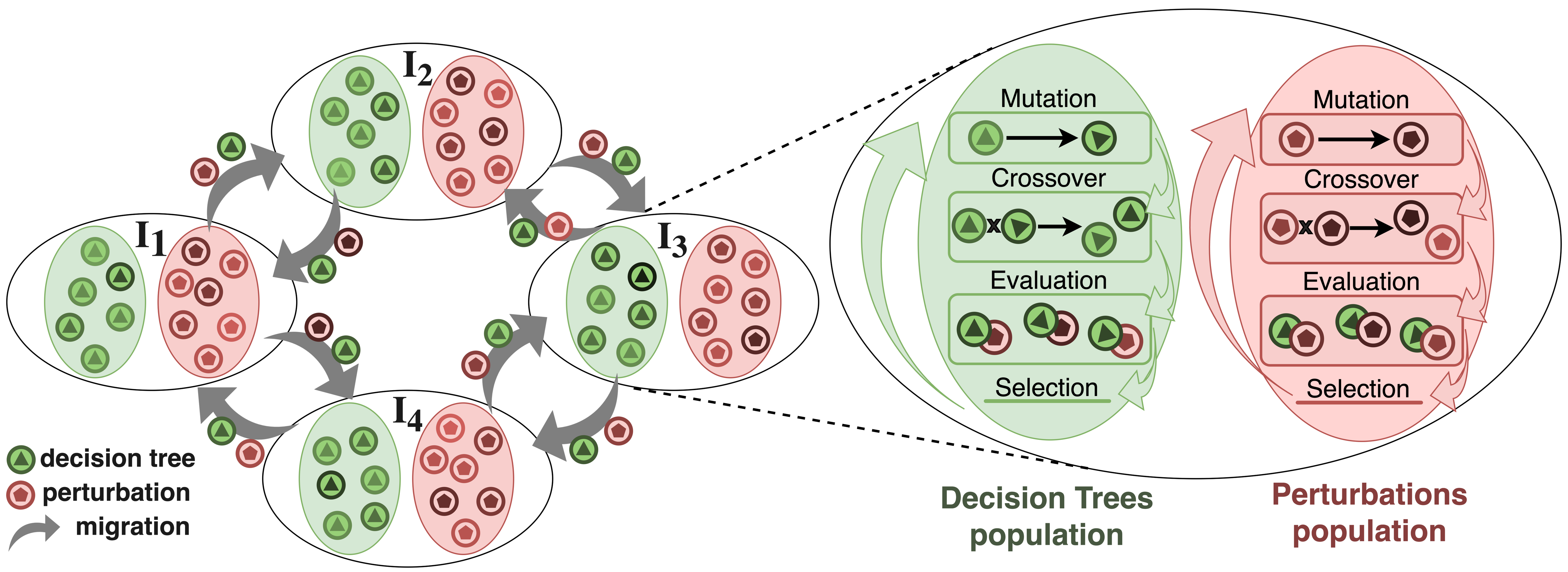
	CS	*I*	*D*
A1	55 (+5)	20 (+10)	0
A2	55 (-5)	20 (+0)	1
A3	55 (-5)	20 (-10)	1



ICoEvoRDF

ICoEvoRDF algorithm leverages island-based coevolution, migration, and game-theoretic principles to enhance the robustness and performance of decision forests.

Algorithm workflow



Each island contains two coevolving populations: Decision Tree (DT) and Perturbation Population. Coevolution alternates between the DT and perturbation populations.

Main results

dataset	Random forests	GROOT forests	FPRDT forest	CoEvoRDT forest	PRAdaBoost	CoEvoRDT boosting	ICoEvoRDF$^{EV}_{SI}$	ICoEvoRDF$_{SI}$	ICoEvoRDF^{EV}	ICoEvoRDF	ICoEvoRDF + FPRDT
ionos	0.112	0.787	0.791	0.793	0.796	0.798	0.797	0.796	0.796	0.799	**0.801**
breast	0.217	0.884	0.873	0.885	0.879	0.899	0.891	0.894	0.896	**0.900**	**0.900**
diabetes	0.452	0.648	0.649	0.621	**0.654**	0.644	0.625	0.636	0.646	0.647	0.651
bank	0.509	0.641	0.658	0.661	0.668	0.669	0.667	0.670	0.664	**0.673**	0.672
Japan3v4	0.519	0.658	0.669	0.679	0.682	0.684	0.684	0.688	0.684	0.688	**0.690**
spam	0.000	0.750	0.749	0.751	0.754	0.763	0.756	0.756	0.762	**0.766**	**0.766**
GesDvP	0.189	0.731	0.725	0.740	0.732	0.753	0.745	0.745	0.749	0.752	**0.754**
har1v2	0.233	0.792	0.828	0.844	**0.860**	0.851	0.855	0.858	0.847	0.854	**0.860**
wine	0.091	0.633	0.681	0.688	0.690	**0.708**	0.691	0.691	0.707	**0.708**	**0.708**
collision-det	0.325	0.726	0.791	0.804	0.800	0.820	0.810	0.812	0.815	**0.822**	**0.822**
mnist-1-5	0.000	0.925	0.964	0.964	0.969	0.975	0.969	0.972	0.968	**0.976**	**0.976**
mnist-2-6	0.000	0.823	0.919	0.917	0.924	0.925	0.923	0.925	0.922	**0.926**	**0.926**
mnist	0.000	0.632	0.750	0.747	0.761	0.763	0.755	0.759	0.759	**0.764**	**0.764**
F-mnist2v5	0.456	0.979	0.974	0.982	0.982	0.993	0.990	0.994	0.987	0.995	**0.996**
F-mnist3v4	0.044	0.839	0.861	0.869	0.867	0.879	0.877	0.877	0.877	**0.884**	**0.884**
F-mnist7v9	0.136	0.836	0.875	0.868	0.879	0.877	0.877	0.880	0.873	**0.881**	0.880
F-mnist	0.024	0.241	0.537	0.545	0.546	0.559	0.552	0.553	0.554	0.560	**0.561**
cifar10:0v5	0.302	0.526	0.683	0.690	0.691	0.699	0.694	0.696	0.697	0.702	**0.703**
cifar10:0v6	0.368	0.560	0.688	0.696	0.696	0.703	0.701	0.701	0.701	0.704	**0.705**
cifar10:4v8	0.296	0.498	0.665	0.665	0.671	0.671	0.674	0.674	0.673	**0.675**	**0.675**
AVERAGE	0.214	0.705	0.767	0.771	0.775	0.782	0.777	0.779	0.779	0.784	**0.785**

Table 1. Averaged adversarial accuracies for ensemble forests methods. The best results are bolded.

Results on **20 datasets** demonstrate the effectiveness of ICoEvoRDF in optimizing both adversarial accuracy and minimax regret metrics. Algorithm consistently **outperforms state-of-the-art methods**, showcasing ability to generate highly robust decision trees and forests. Use of island-based coevolution and game-theoretic weighting strategies proved particularly advantageous, **improving diversity** and **leading to more robust decision tree ensembles**.

Algorithm details

Initialization: Unique training sets assigned to each island sampled with replacement from dataset.

Evolutionary operators: mutation, crossover, selection.

Evaluation: Each population is evaluated against individuals from the opposing population.

Migration: Introduces genetic diversity by sharing solutions between neighboring islands based on an island topology (e.g. ring topology).

Decision Forest Composition: The final decision forest is constructed using **the fittest DTs from all islands** with weighted voting:
Equal voting (EV): the same contribution from each island representative (basic approach).
Nash-Based Voting (NV): Frame the scenario as a **two-player game**: DT player chooses strategies from the fittest DTs, perturbation player chooses strategies from perturbations. **Use mixed Nash equilibrium probabilities as voting weights.**

Conclusions

- Independently **evolving populations of decision trees and perturbations**, with **periodic migration of top-performing individuals** between islands.
- ICoEvoRDT **fosters diversity** and promotes the **exploration of a wider range of potential solutions**.
- **Synergy between coevolutionary methods and game theory** (Nash equilibrium based voting).
- **Trade-off between model interpretability and robustness** - ICoEvoRDF can produce more robust ensemble models or easier to interpret single DTs.

Scan for arXiv paper

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