



Contribution ID: 101

Type: not specified

Agreement over Association in Explainable Ensemble Trees

Ensemble methods such as Random Forests achieve strong predictive accuracy but at the cost of interpretability. Explainable Ensemble Trees (E2Tree) address this trade-off by constructing a single interpretable tree that approximates the co-occurrence structure induced by the ensemble. The quality of this approximation matters: when interpretability is invoked for regulatory or scientific purposes, a poor reconstruction does not merely underperform: it actively misleads.

Existing validation approaches for E2Tree rely on the Mantel test, which measures the association between proximity matrices. We argue that this answers the wrong question. The issue parallels the classical distinction between correlation and concordance in method comparison: two matrices can be perfectly correlated while differing substantially in absolute terms. What E2Tree validation requires is a scale-sensitive measure of agreement.

We propose a family of divergence and similarity measures for this task. The centrepiece is the Normalized Loss of Interpretability (nLoI), a statistic rooted in the Cressie–Read power divergence family, whose key feature is a decomposition into within-node and between-node components. This identifies not only how much reconstruction quality is lost, but where and why, a diagnostic capability unavailable from correlation-based approaches. Four complementary measures complete the family: Hellinger distance, weighted Root Mean Squared Error, the RV coefficient, and the Structural Similarity Index, each targeting a distinct facet of matrix agreement.

A unified permutation testing framework based on simultaneous row/column permutation provides valid inference for all measures. Monte Carlo simulations confirm correct Type-I error control and adequate power; empirical results on benchmark datasets illustrate the framework's utility.

Special/ Invited session

ISBIS

Classification

Mainly methodology

Keywords

Ensemble Models, Interpretability, Agreement Measures

Primary author: GNASSO, Agostino (University of Naples Federico II)

Co-authors: IORIO, Carmela (University of Naples Federico II); Prof. ARIA, Massimo (University of Naples Federico II)

Presenter: GNASSO, Agostino (University of Naples Federico II)

Track Classification: Other/special session/invited session