Error Detection and Constraint Recovery in Hierarchical Multi-Label Classification without Prior Knowledge

Joshua Shay Kricheli, Khoa Vo, Aniruddha Datta, Spencer Ozgur, Paulo Shakarian
Arizona State University

Abstract

Recent advances in Hierarchical Multi-label Classification (HMC), particularly neurosymbolic-based approaches, have demonstrated improved consistency and accuracy by enforcing constraints on a neural model during training. However, such work assumes the existence of such constraints a-priori. In this paper, we relax this assumption and present an approach based on Error Detection Rules (EDR) that allow for learning explainable rules about the failure modes of machine learning models. We show that these rules are not only effective in detecting when a machine learning classifier has made an error but also can be leveraged as constraints for HMC, thereby allowing the recovery of explainable constraints even if they are not provided. We show that our approach is effective in detecting machine learning errors and recovering constraints, is noise tolerant, and can function as a source of knowledge for neurosymbolic models on multiple datasets, including a newly introduced military vehicle recognition dataset.

Motivation and Contribution

- **Hierarchical Multi-label Classification (HMC)** extends the idea of multi-label classification to impose a hierarchy among labels
- Some recent research trends have led to the expression of hierarchical relationships as constraints on the learning process, usually requiring prior knowledge of the constraints
- **Metacognition** deals with employing an auxiliary learning model for reasoning tasks about an existing model, such as error detection and correction
- In this work we extend the metacognitive EDR framework which detects and corrects errors of a learning model, to address the HMC problem without prior knowledge of the hierarchy constraints
- We further present the novel Focused-EDR extension to EDR, which proposes an improved objective function for better overall performance of the meta-model in detecting errors of the original model

Focused Error Detection Rules (f-EDR)

- Error Detection Rules (EDR) is a metacognitive approach for detecting errors in the result of a trained machine learning model that assigns a label y for some sample x
- The method utilizes a set of boolean conditions C, associated with each sample and assignable from domain knowledge or a complementary model for the same task
- The key intuition is for each class y to identify a subset of conditions $DC_y \subseteq C$ to form an error detection rule:

$$error_y(X) \leftarrow assign_y(X) \wedge \bigvee_{cond \in DC_y} cond(X)$$

- In the novel f-EDR approach, DC_y is constructed using a greedy algorithm that approximates a solution which maximizes the f-1 score of the class y after applying the rule, instead of its precision, as the EDR method does

Results

- SOTA models were evaluated in this work on three datasets– a Military Vehicles dataset, which we scraped and published, and subsets of 50 classes from the ImageNet dataset and 36 classes from the OpenImage dataset
- Each dataset has two levels of labels in its hierarchy: fine and coarse grain
- For error detection (Table 1), we evaluated three approaches: f-EDR, EDR, and a binary neural network for error prediction, inspired by related work
- We compared them using the average balanced error accuracy and f1 score across all classes
- We proceeded to use these rules to perform error correction (Table 2) by adding the learned rules as constraints to the loss function, using the Logic Tensor Networks (LTN) method, which we compared with the baseline with the accuracies for each level of the hierarchy and the inconsistenices among them
- We also conducted noise tolerance experiments (Figure 1), in which we added noise to a fraction of the classes at training and evaluated the accuracy and constrains f1 score

Table 1: Error Detection Results for all the datasets

Dataset	Method	Balanced Error Acc.	Error f1
Military Vehicles	Binary NN	80.10%	80.18%
	EDR	83.45%	82.62%
	f-EDR (ours)	**84.08%**	**83.17%**
ImageNet50	Binary NN	72.85%	68.96%
	EDR	80.92%	72.78%
	f-EDR (ours)	**84.26%**	**77.78%**
OpenImage36	Binary NN	64.80%	63.65%
	EDR	59.87%	46.46%
	f-EDR (ours)	**66.63%**	**65.83%**

Table 2: Error Correction Results with LTN

Dataset	Method	Fine-Grain Acc.	Fine-Grain f1	Coarse-Grain Acc.	Coarse-Grain f1	Inconsistency
Military Vehicles	VIT b_16	54.35%	48.40%	77.24%	74.57%	7.77% (126/1621)
	f-EDR + LTN (ours)	**62.43%**	**58.18%**	**82.17%**	**79.95%**	**5.37%** (87/1621)
ImageNet50	DINO V2 s_14	85.76%	85.59%	93.52%	92.65%	1.19% (25/2100)
	f-EDR + LTN (ours)	**86.29%**	**86.12%**	**93.57%**	**92.69%**	**1.05%** (22/2100)
OpenImage36	VIT b_16	57.68%	55.89%	90.15%	88.82%	3.02% (362/12002)
	f-EDR + LTN (ours)	**60.11%**	**58.88%**	**91.21%**	**89.85%**	**1.73%** (208/12002)

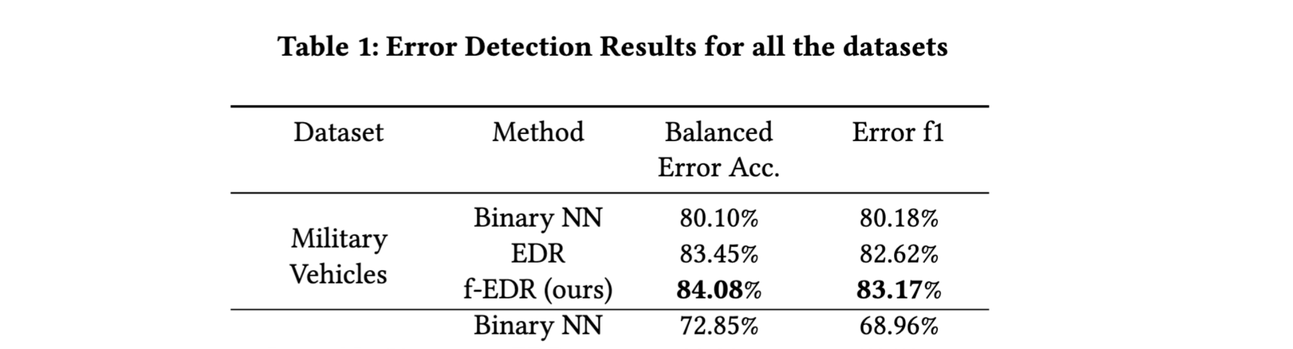

Figure 1: Balanced Error Accuracy, Error F1-score, and constraints F1-score results of the Focused-EDR method for varying noise ratios for the Military Vehicles (a) ImageNet50 (b) and OpenImage36 (c) datasets

V2 LAB

ASU Ira A. Fulton Schools of Engineering
ARIZONA STATE UNIVERSITY

CIKM 2024 OCTOBER 21-25

BOISE

Visit https://neurosymbolic.asu.edu for more information