Accelerating Constraint-Based Neural Network Repairsby Example Prioritization and Selection

Long ZHANG, Shuo SUN, Jun YAN, Jian ZHANG, Jiangzhao WU, Jian LIU

Frontiers of Computer Science, DOI: 10.1007/s11704-024-3902-x

Problems & Ideas

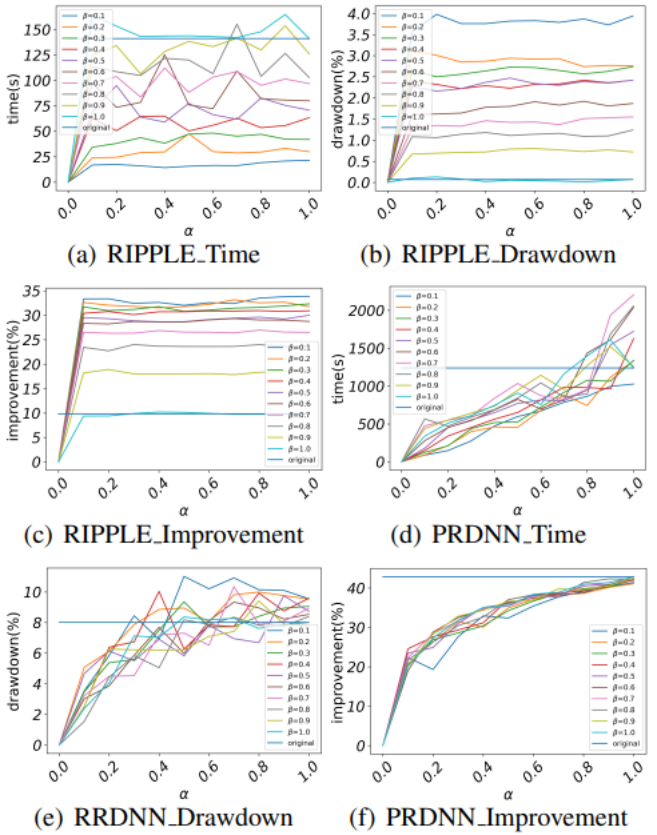
- Problems of Constraint-Based Neural Network Repairs:
  - Computational efficiency can be adversely affected when the example set encompasses excessive examples, resulting in calculation explosions, prolonged execution times, or out-of-memory errors.
  - The example set can be reduced to boosting constraint-based neural network repairs, but determining which examples within the example set can be minimized to expedite neural network repair remains a formidable task.
- Ideas: we observed that not all examples are equally useful for the repair process, therefore we can identify the examples of less value and thereby remove the redundant constraints in the models. Example prioritization and selection can be used to reduce the size of example set.

Main Contributions

- Contributions:
  - We define a metric to estimate the confidence level of the neural network's predictions on examples, and propose an example selection strategy to filter out examples for the neural network repairs.
  - we present a novel approach that incorporates example prioritization and selection strategies to accelerate the neural network repair process.

Table 1 The *Drawdown*, *Improvement*, and *Time* effectiveness of our approach on different datasets and repair approaches

			Boosted	Original
PRDNN	f-MNIST	Drawdown(%)	**3.09**	4.85
		Improvement(%)	18.57	23.02
		Time(s)	**391.12**	1282.99
	MNIST	Drawdown(%)	**2.85**	7.95
		Improvement(%)	34.23	42.82
		Time(s)	**774.48**	1223.89
	CIFAR10	Drawdown(%)	**2.09**	–
		Improvement(%)	**2.48**	–
		Time(s)	**1,621**	–
RIPPLE	f-MNIST	Drawdown(%)	0.46	0.14
		Improvement(%)	**16.67**	15.44
		Time(s)	**447.28**	711.26
	MNIST	Drawdown(%)	2.65	0.07
		Improvement(%)	**31.61**	9.85
		Time(s)	**48.44**	130.94
	CIFAR10	Drawdown(%)	0.32	0.28
		Improvement(%)	**2.31**	2.05
		Time(s)	**816.71**	1672.76



(a) RIPPLE_Time (b) RIPPLE_Drawdown
(c) RIPPLE_Improvement (d) PRDNN_Time
(e) RRDNN_Drawdown (f) PRDNN_Improvement

Fig. 1 The effect of hyperparameters α and β