

Interpretable Classification Rules in Relaxed Logical Form

Bishwamittra Ghosh

Joint work with
Dmitry Malioutov and Kuldeep S. Meel



Machine learning algorithms continue to permeate critical application domains

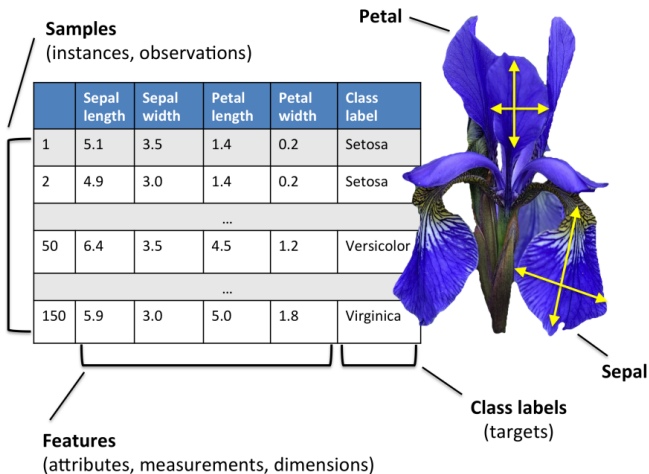
- ▶ medicine
- ▶ legal
- ▶ transportation
- ▶ ...

It becomes increasingly important to

- ▶ understand ML decisions

Interpretability has become a central thread in ML research

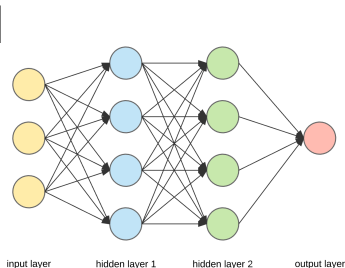
Example Dataset



Representation of an interpretable model and a black box model

A sample is **Iris Versicolor** if
(sepal length > 6.3 **OR** sepal width > 3
OR petal width ≤ 1.5)
AND
(sepal width ≤ 2.7 **OR** petal length > 4
OR petal width > 1.2)
AND
(petal length ≤ 5)

Interpretable Model



Black Box Model

CNF Formula

- ▶ A **CNF** (Conjunctive Normal Form) formula is a **conjunction** of clauses where each clause is a **disjunction** of literals
- ▶ A DNF (Disjunctive Normal Form) formula is a disjunction of clauses where each clause is a conjunction of literals
- ▶ Example
 - ▶ CNF: $(a \vee \neg b \vee c) \wedge (\neg d \vee e)$
 - ▶ DNF: $(a \wedge b \wedge \neg c) \vee (\neg d \wedge e)$
- ▶ Decision rules in CNF and DNF are highly interpretable [Malioutov'18; Lakkaraju'19]

Definition of Interpretability in Rule-based Classification

- ▶ There exists different notions of interpretability of rules
- ▶ Rules with **fewer terms** are considered interpretable in medical domains [Letham'15]

$$\mathcal{R} = (a \vee b \vee \neg c \vee d \vee e) \wedge \\ (f \vee g \vee h \vee \neg i) \wedge \\ (j \vee k \vee \neg l) \wedge \\ (\neg m \vee n \vee o \vee p \vee q) \wedge$$

$$\mathcal{R} = (a \vee b \vee \neg c) \wedge \\ (f \vee g)$$

- ▶ We consider **rule size** as a proxy of interpretability for rule-based classifiers
- ▶ For CNF/DNF, rule size = number of literals

Outline

Introduction

Example of Interpretable Rules

Motivation

Formulation of relaxed-CNF

Experiments

Future Work and Conclusion

Can we design a classifier to generate a richer family of logical rules?

Our Contribution

- ▶ generalize the widely popular CNF rules and introduce a richer family of logical rules
 - ▶ introduce **relaxed-CNF** rules
- ▶ propose a scalable framework for learning relaxed-CNF rules

CNF

$$(a \vee \neg b \vee c) \wedge (\neg d \vee e \vee f)$$

CNF

$$[(a + \neg b + c) \geq 1] + [(\neg d + e + f) \geq 1] \geq 2$$

Relaxed-CNF

$$[(a + \neg b + c) \geq \eta_l] + [(\neg d + e + f) \geq \eta_l] \geq \eta_c$$

$$0 \leq \eta_l \leq \text{number of literals}$$

$$0 \leq \eta_c \leq \text{number of clauses}$$

Definition of Relaxed-CNF

- ▶ Relaxed-CNF formula has two parameters η_l and η_c
- ▶ A clause is satisfied if at least η_l literals are satisfied
- ▶ A formula is satisfied if at least η_c clauses are satisfied

Applications



Figure: Checklist

CHADS₂ risk criteria	Score
Congestive heart failure	1
Hypertension	1
Age > 75	1
Diabetes mellitus	1
Prior stroke or transient ischemic attack	2
CHADS₂ score	Adjusted stroke rate (% per year)
Stroke risk based on CHADS₂ score	
0	1.9%
1	2.8%
2	4.0%
3	5.9%
4	8.5%
5	12.5%
6	18.2%

Figure: Stroke prediction in medical domain

Benefit of Relaxed-CNF form

- ▶ Relaxed-CNF is more succinct than CNF
- ▶ Rule size = number of literals

$$\underbrace{(a + b + c) \geq 2}_{\text{rule size: 3}} \Rightarrow \underbrace{(a \vee b) \wedge (a \vee c) \wedge (b \vee c)}_{\text{rule size: 6}}$$

A single clause in relaxed-CNF is equivalent to exponential number of clauses in CNF

IRR: Interpretable Rules in Relaxed Form

- ▶ We design objective function to
 - ▶ minimize prediction error
 - ▶ minimize rule size (i.e. maximize interpretability)
 - ▶ feature variable: $b = \mathbb{1}\{\text{feature is selected in rule}\}$
 - ▶ noise variable: $\xi = \mathbb{1}\{\text{sample is misclassified}\}$

$$\min \sum \xi + \lambda \sum b$$

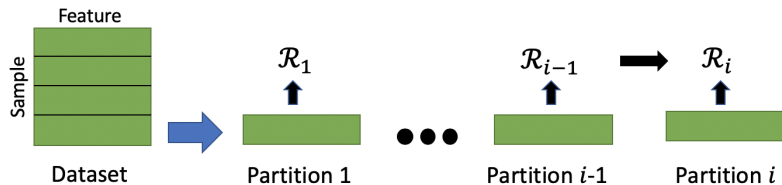
IRR: Interpretable Rules in Relaxed Form

- ▶ We design objective function to
 - ▶ minimize prediction error
 - ▶ minimize rule size (i.e. maximize interpretability)
 - ▶ feature variable: $b = \mathbb{1}\{\text{feature is selected in rule}\}$
 - ▶ noise variable: $\xi = \mathbb{1}\{\text{sample is misclassified}\}$

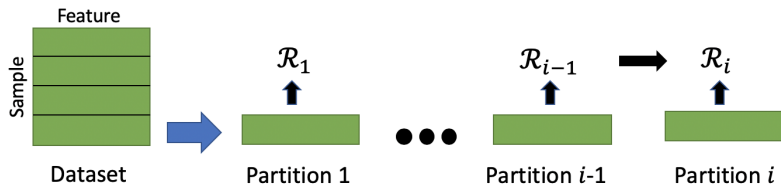
$$\min \sum \xi + \lambda \sum b$$

- ▶ We formulate an Integer Linear Program (ILP) for learning relaxed-CNF rules
- ▶ We incorporate incremental learning in ILP formulation to achieve scalability

Incremental Approach



Incremental Approach



Modified objective function:

$$\min \sum \xi + \lambda \sum b \cdot \mathbb{I}(b)$$

where

$$\mathbb{I}(b) = \begin{cases} -1 & \text{if } b = 1 \text{ in previous partition} \\ 1 & \text{otherwise} \end{cases}$$

Experimental Results

Accuracy of relaxed-CNF rules and other classifiers

Dataset	size	feature	SVC	RF	RIPPER	IMLI	IRR	inc-IRR
Heart	303	31	85.48	83.87	81.59	80.65	86.65	86.44
WDBC	569	88	98.23	96.49	96.49	96.46	97.34	96.49
TicTacToe	958	27	98.44	99.47	98.44	82.72	84.37	84.46
Titanic	1309	26	78.54	79.01	78.63	79.01	81.22	78.63
Tom's HW	28179	910	97.6	97.46	97.6	96.01	97.34	96.52
Credit	30000	110	82.17	82.12	82.13	81.75	82.15	81.94
Adult	32561	144	87.19	86.98	84.89	83.63	85.23	83.14
Twitter	49999	1511	—	96.48	96.14	94.57	95.44	93.22

Table: Test accuracy (%) of different classifiers.

Summary:

- ▶ IRR has competitive accuracy compared to other classifiers
- ▶ IRR times out in most datasets
- ▶ inc-IRR achieves scalability with a little loss of accuracy

Rule-size of different interpretable models

Dataset	RIPPER	IMLI	inc-IRR
Heart	7	14	19.5
WDBC	7	11	10
Tic Tac Toe	25	11.5	12
Titanic	5	7	12.5
Tom's HW	16.5	32	5.5
Credit	33	9	3
Adult	106	35.5	13
Twitter	56	67.5	7

Table: Rule size of different interpretable classifiers.

Summary:

- For larger datasets, rule size of relaxed-CNF is smaller

Conclusion

- ▶ Relaxed-CNF rules allow increased flexibility to fit data
- ▶ The size of relaxed-CNF rule is less for larger datasets, indicating higher interpretability
- ▶ Smaller relaxed-CNF rules reach the same level of accuracy compared to plain CNF/DNF rules and decision lists

Future Works

- ▶ Human evaluations of relaxed-CNF
- ▶ More scalable and robust design of framework by adopting ILP techniques: column generation, lp-relaxation etc.
- ▶ Calculating the capacity of relaxed-CNF using VC dimension

Source code: <https://github.com/meelgroup/IRR>

Thank You