Incremental Approach to Interpretable Classification Rule Learning

Bishwamittra Ghosh and Kuldeep S. Meel
School of Computing, National University of Singapore

CP 2019

Practical applications of machine learning

- Hiring employees
- Giving a loan to a person
- Predicting recidivism: likelihood of a person convicted of a crime to offend again
- ...

Practical applications of machine learning

- Hiring employees
- Giving a loan to a person
- Predicting recidivism: likelihood of a person convicted of a crime to offend again
- . . .

Should we believe the prediction of machine learning models?

Practical applications of machine learning

- Hiring employees
- Giving a loan to a person
- Predicting recidivism: likelihood of a person convicted of a crime to offend again
- ...

Should we believe the prediction of machine learning models?

Interpretable classification model

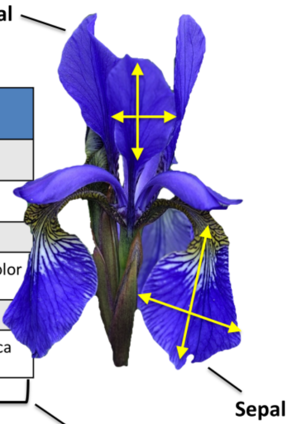
Example Dataset

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			...		
50	6.4	3.5	4.5	1.2	Versicolor
			...		
150	5.9	3.0	5.0	1.8	Virginica

Features

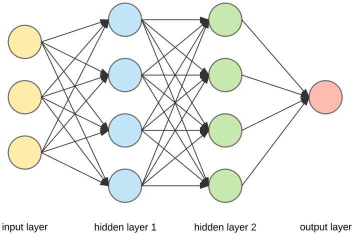
Class labels
(targets)



Representation of an interpretable model and a black box model

A sample is predicted as 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

Formula

- A CNF (Conjunctive Normal Form) formula is a conjunction of clauses where each clause is a disjunction of literals

$$(a \lor \neg b \lor c) \land (d \lor e)$$

- A DNF (Disjunctive Normal Form) formula is a disjunction of clauses where each clause is a conjunction of literals

$$(a \land b \land \neg c) \lor (d \land e)$$

Formula

- A CNF (Conjunctive Normal Form) formula is a conjunction of clauses where each clause is a disjunction of literals

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

- A DNF (Disjunctive Normal Form) formula is a disjunction of clauses where each clause is a conjunction of literals

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

- Decision rules in CNF and DNF are highly interpretable [Malioutov'18; Lakkaraju'19]

Definition of interpretability in rule-based classifiers

- There exists different notions of interpretability of rules

Definition of interpretability in rule-based classifiers

- There exists different notions of interpretability of rules

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

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

- Rules with fewer terms are considered interpretable in medical domains [Letham'15]

Definition of interpretability in rule-based classifiers

- There exists different notions of interpretability of rules

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

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

- Rules with fewer terms are considered interpretable in medical domains [Letham'15]
- We refer rule size as a proxy of interpretability in rule-based classifiers
- For rules expressed as CNF/DNF, rule size = number of literals

Outline

Design of an interpretable classifier [Malioutov'18]

- We design objective function to
  - minimize prediction error
  - minimize rule size (i.e., maximize interpretability)

Design of an interpretable classifier [Malioutov'18]

- We design objective function to
  - minimize prediction error
  - minimize rule size (i.e., maximize interpretability)
- Consider decision variables:
  - feature variables $b_i^j = \mathbb{1}\{j\text{-th feature is selected in } i\text{-th clause}\}$
  - noise variables $\eta_q = \mathbb{1}\{\text{sample } q \text{ is misclassified}\}$

$$\min \sum_{i,j} b_i^j + \lambda \sum_q \eta_q$$

- Constraints:
  - a positive labeled sample satisfies the rule
  - a negative labeled sample does not satisfy the rule
  - otherwise the sample is considered as noise

MaxSAT

In MaxSAT

- **Hard Clause:** always satisfied, weight $= \infty$
- **Soft Clause:** can be falsified, weight $= \mathbb{R}^+$

MaxSAT finds an assignment that satisfies all hard clauses and most soft clauses such that the weight of satisfied soft clauses is maximized

MaxSAT-based approach for interpretable rule-based classification

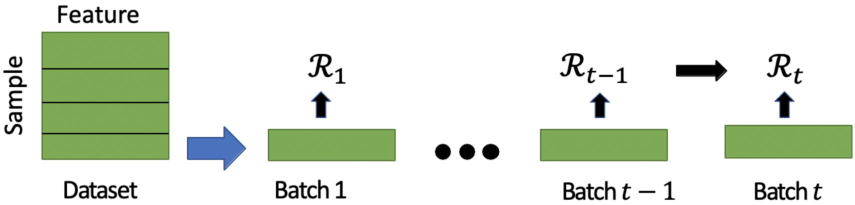
- the objective function is encoded as soft clauses
- the constraints are encoded as hard clauses

Analysis

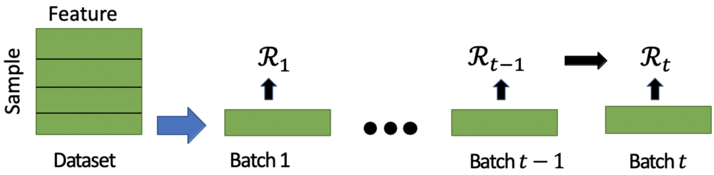
- To generate a k-clause CNF rule for a dataset of n samples over m boolean features, the number of clauses of the MaxSAT instance is $\mathcal{O}(n \cdot m \cdot k)$

An Incremental Rule-learning Approach [Ghosh'19]

- We attribute large formula size of the MaxSAT instance for the poor scalability
- We propose mini-batch incremental learning



Solution Technique



- We propose a mini-batch incremental learning framework with the following objective function on batch t

$$\min \sum_{i,j} b_i^j \cdot I(b_i^j) + \lambda \sum_q \eta_q.$$

where indicator function $I(\cdot)$ is defined as follows.

$$I(b_i^j) = \begin{cases} -1 & \text{if } b_i^j = 1 \text{ in the } (t-1)\text{-th batch } (t \neq 1) \\ 1 & \text{otherwise} \end{cases}$$

Continued...

$(t-1)$-th batch

we learn assignment

- $b_1 = 0$
- $b_2 = 1$
- $b_3 = 0$
- $b_4 = 1$

t-th batch

we construct soft unit clause

- $\neg b_1$
- b_2
- $\neg b_3$
- b_4

Experimental Results

Accuracy and training time of different classifiers

Dataset	Size n	Features m	LR	SVC	RIPPER	IMLI
PIMA	768	134	75.32 (0.3s)	75.32 (0.37s)	75.32 (2.58s)	73.38 (0.74s)
Credit-default	30000	334	80.81 (6.87s)	80.69 (847.93s)	80.97 (20.37s)	79.41 (32.58s)
Twitter	49999	1050	95.67 (3.99s)	Timeout	95.56 (98.21s)	94.69 (59.67s)

Table: Each cell in the last 5 columns refers to test accuracy (%) and training time (s).

IMLI exhibits better training time by costing a little bit of accuracy

Size of rules of different rule-based classifiers

Dataset	RIPPER	IMLI
PIMA	8.25	3.5
Twitter	21.6	6
Credit	14.25	3

Table: Average size of the rules of different rule-based models.

IMLI generates shorter rules compared to other rule-based models

Conclusion

- Interpretable ML model ensures reliability of prediction models in practice
- We propose an incremental learning approach of classification rules
- IMLI[1] achieves up to three orders of magnitude improvement in training time by sacrificing a bit of accuracy
- The generated rules appear to be more interpretable

Python library:

```
$ pip install rulelearning
```

Thank You !!

[1] Source code: `https://github.com/meelgroup/MLIC`