Constructing Decision Rules from Naive Bayes Model for Robust and Low Complexity Classification

Nabeel Al-A'araj **Safaa O. Al-mamory** **Ali H. Al-Shakarchi**

Global Research Conference (GRaCe 2020)

Agenda

- Introduction
- Motivation
- Contribution
- Rule Extraction Approaches
- RULE Extraction Methodology
- Experiments
- Results
- Conclusion

Introduction

- Classification is a vital branch of data mining which is utilized by many applications
- DT, NN, SVM, RBC, and BC are used for classification and pattern recognition tasks in many enterprises and industries
- Classification techniques can be grouped according to two major criteria

How the model has been built

- DT model is derived and expressed as a tree
- RBC model is derived and expressed as (IF TEHN rules)

The complexity of the model

- Some classifiers provide explanation capability for decision-making process like (DT, RBC), while other produce black-box models such as(SVM, NN

Motivation

- NB is a probabilistic model. This model makes its prediction depend on probability membership.

- NB has two major limitations:
  A. Its time consuming if it applied on a large dataset.(the whole dataset needs to be scanned to apply statistical equations that perform classification.)
  B. NB may remain difficult for non-statisticians.(understand the detail operation and classification mechanism of a model, thus, it could be considered as a black-box model from the perspective of those who are not specialists in statistics.)

- RBC relays on its prediction on a rule-set that is more comprehensible for human.

IF (Condition) → THEN → **Conclusion**

If we exploit the pros of both techniques (i.e. NB, RBC), we can get the classification power of NB and elevate its limitation

Rule Extraction Approaches

RULE Extraction Methodology

- To construct rules, the straightforward way is to construct a rule for every attribute of patterns.

```
Algorithm: Rule Constructing from NB
Input: NB conditional probability
       p(X_i|C), threshold T
Output: Rule-set (R)
  Begin
      for each instance X_i do
        for each value of class C do
         P: = p(X_i|C_k)
            if (P>=T) OR (P<=1-T) then
           R=R U(X_i -->C_k:f(p))
        end
      end
   End
```

RULE Extraction Methodology

- What ideal RBC should satisfy?

  1. **Mutually exclusive (**there are no two rules triggered by the same pattern **).**
  2. **Exhaustive (**there is only one rule for every combination of feature values**)**

ensure that every example is covered exactly by one rule

- There are no ideal RBCs, and this motivates researchers to find an exir to resolve conflict rules.

- Our implementation of classification rules will adopt two different strategies. They are

  A)class-based ordering and

  B) size-ordering

- Equal width(EW) and Equal frequency(EF) are two unsupervised binning methods that have been used in the discretization process.

Experiments

- Three different datasets have been selected: Wisconsin Breast Cancer (WBC), Vote, and Diabetes datasets .

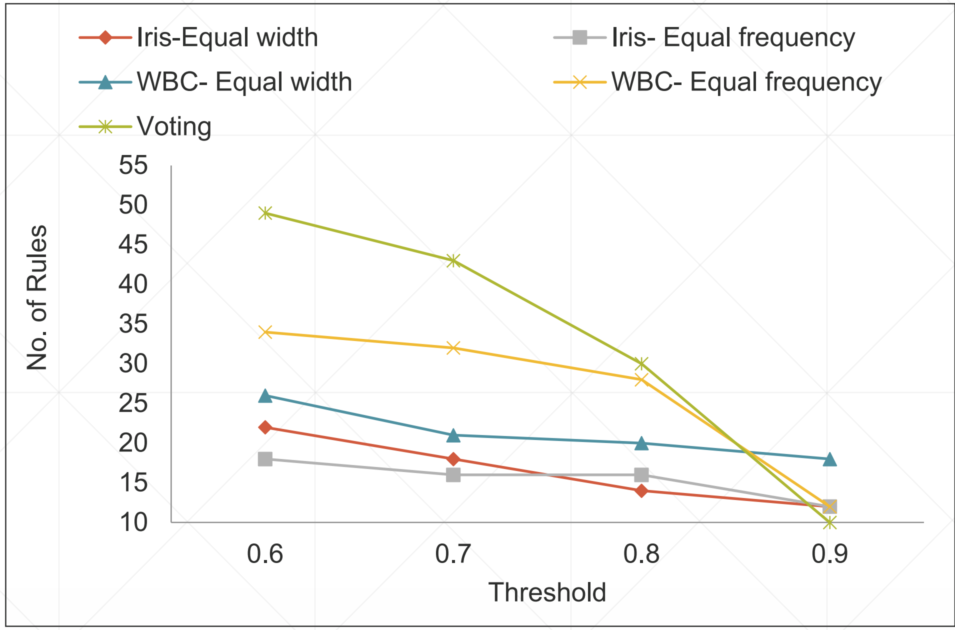
Dataset	No. of classes	No. of attributes	Missing value	No. of nominal attribute	No. of numerical attribute	No. of instances
WBC	2	9	Several	0	9	699
Voting	2	16	Many	16	0	435
Diabetes	2	9	None	0	9	768

Results

Number of Constructed Rules for Dataset (Without Default Rules)

T	Iris , bin=2		WBC , bin=2		Voting
	EW	EF	EW	EF	
0.6	22	18	26	34	49
0.7	18	16	21	32	43
0.8	14	16	20	28	30
0.9	12	12	18	12	10
T	Iris , bin=3		WBC , bin=3		
	EW	EF	EW	EF	
0.6	33	35	49	48	
0.7	30	31	39	41	
0.8	26	25	30	33	
0.9	24	24	27	18	

No. of Generated Rules Per Threshold T With Bin=2

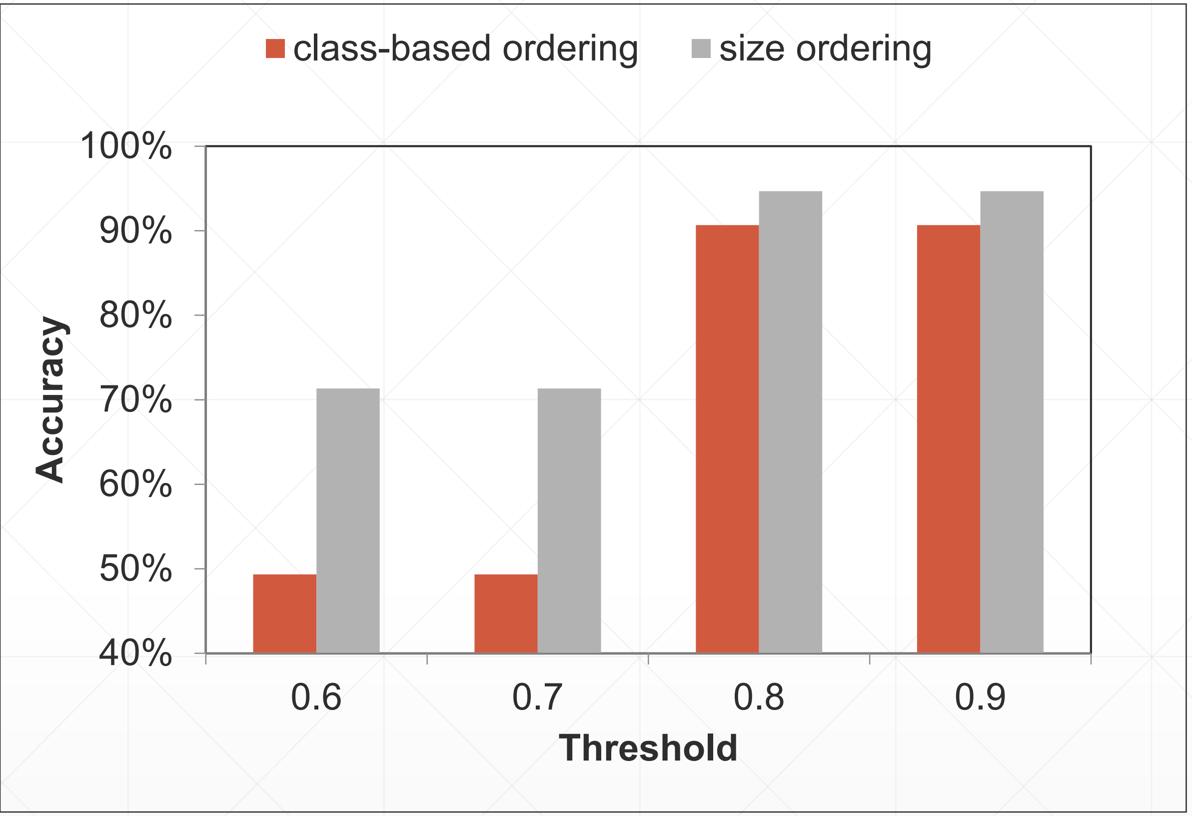


Results

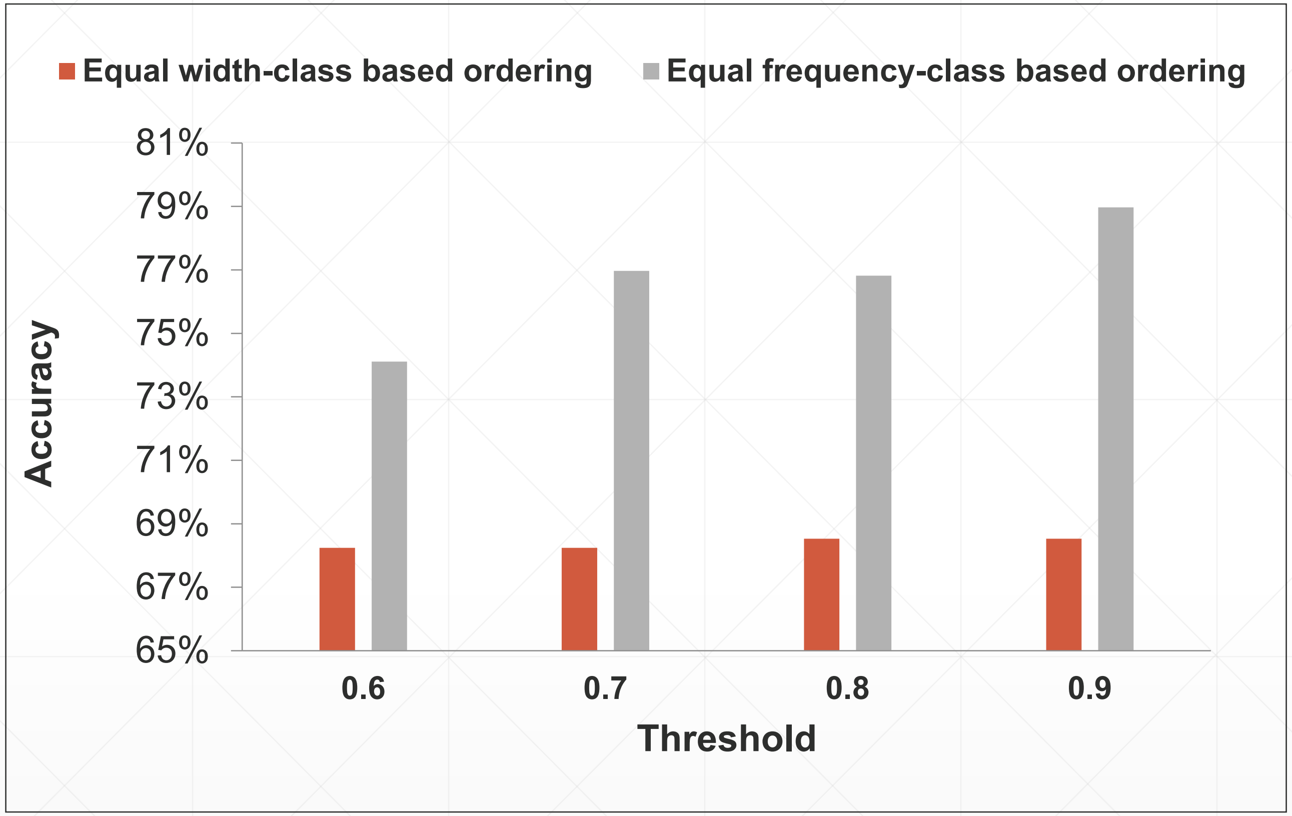
Accuracy of Generated Rule-Set for (Iris, WBC) in Different Configuration (A) class-based ordering and (B) size-ordering

Iris Dataset				
T	Binning	# Rules	Conflict Strategy	Accuracy
0.6	EW	10	A	49.33%
			B	71.33%
	EF	10	A	74.67%
			B	78.67%
0.7	EW	9	A	49.33%
			B	71.33%
	EF	9	A	80.0%
			B	80.0%
0.8	EW	7	A	90.67%
			B	94.67%
	EF	7	A	91.33%
			B	95.33%
0.9	EW	6	A	90.67%
			B	94.67%
	EF	7	A	91.33%
			B	95.33%
WBC Dataset				
T	Binning	# Rules	Conflict strategy	Accuracy
0.6	EW	11	A	68.24%
			B	68.24%
	EF	12	A	74.11%
			B	74.11%
0.7	EW	10	A	68.24%
			B	68.24%
	EF	11	A	76.97%
			B	76.97%
0.8	EW	9	A	68.53%
			B	68.53%
	EF	8	A	76.82%
			B	76.82%
0.9	EW	8	A	68.53%
			B	68.53%
	EF	2	A	78.97%
			B	78.97%

Results



Accuracy Comparison Between Class-Based Ordering And Size Ordering With EW Binning on Iris Dataset



Accuracy Comparison Between Equal Width And Equal Frequency on WBC Dataset

Results

Evaluation of Iris Rule-Set

Rules	Coverage	Accuracy	Laplace	M_Estimate
if (sepallength<-5.5) =====>Iris-setosa:1.0	34.67%	86.54%	83.64%	81.82%
if (petallength<2.5) =====>Iris-setosa:1.0	33.33%	100.0%	96.23%	94.34%
if (petalwidth<0.8) =====>Iris-setosa:1.0	33.33%	100.0%	96.23%	94.34%
if (petallength>=2.5 && petallength <4.8) =====>Iris-versicolor:1.0	30.0%	97.78%	93.75%	91.67%
if (petalwidth>=0.8&& petalwidth <1.6) =====>Iris-versicolor:1.0	32.0%	93.75%	90.2%	88.24%
if (petallength>=4.8) =====>Iris-virginica:1.0	36.67%	89.09%	86.21%	84.48%
if (petalwidth>=1.6)=====>Iris-virginica:1.0	34.67%	34.67%	87.27%	85.45%

Conclusions

- The experimental results indicate
  - ❖ Rule set constructed from Naïve Bayes is pure with classification accuracy relatively high. In addition , in some datasets, the constructed rule set shows better accuracy compared with original NB model such as in (Voting).
  - ❖ That EF gave better accuracy of rule set than EW.

THANK YOU

Questions

?