

Random Rules From Data Streams

Ezilda Almeida
LIAAD - INESC Porto L.A.,
Portugal
ezildacv@gmail.com

Petr Kosina
LIAAD - INESC Porto L.A.,
Portugal
and Fac. of Informatics,
Masaryk University, Brno
petr.kosina@inescporto.pt

João Gama
LIAAD - INESC Porto L.A.,
Portugal
and Fac. of Economics,
University of Porto, Portugal
jgama@fep.up.pt

ABSTRACT

Existing works suggest that random inputs and random features produce good results in classification. In this paper we study the problem of generating random rule sets from data streams. One of the most interpretable and flexible models for data stream mining prediction tasks is the Very Fast Decision Rules learner (VFDR). In this work we extend the VFDR algorithm using random rules from data streams. The proposed algorithm generates several sets of rules. Each rule set is associated with a set of N_{att} attributes. The proposed algorithm maintains all properties required when learning from stationary data streams: online and any-time classification, processing each example once.

Keywords

Data Streams, Classification, Rule Learning, Random Rules

Categories and Subject Descriptors

I.2.6 [Artificial Intelligence]: Learning—*Concept learning*

1. MOTIVATION

Important recent problems, such as medical diagnosis and document retrieval, are often composed of many input variables. It is not unlikely that the number of variables is in the hundreds or thousands, with each one containing only a small amount of information. Consequently a single rule set might not be able to achieve high performance and there might be a space for improvement. There were significant improvements in classification accuracy reported in the literature by using random inputs and random features. Therefore combining sets of rules using random features can improve the performance.

In the work presented in [3] the authors introduced VFDR. In this paper we present random rules from data streams (RR) which is an extension of the rule set classifier for data streams VFDR.

2. RELATED WORK

In this section we briefly present some of the research literature related to data streams, random forests and random prism.

Decision rules from data streams. In [3] the authors present the Very Fast Decision Rules algorithm. VFDR is designed for high-speed data streams. It is a single pass algorithm that learns ordered and/or unordered rules.

Random forests. This method was proposed in [1], and consists of a collection or ensemble of simple tree predictors, each capable of producing a response when presented with a set of predictor values. To determine the class of an instance, the method combines the result of various decision trees using a voting mechanism.

Random prism. In [5] the authors present the Random Prism algorithm. Random Prism classifier is inspired from the Prism family of algorithms [2], the Random Decision Forests and Random Forests approaches. Random Prism uses the PrismTCS (Target Class Smallest first) [4] classifier as base classifier with some modifications called R-PrismTCS.

3. THE RANDOM RULES ALGORITHM

The algorithm's main parameters are the number of rule sets (N_{rs}) and the number of attributes (N_{att}), where N_{att} must be less than the total number of attributes (M). Initially each rule set is empty. Each rule set is associated with a subset of N_{att} attributes that are selected randomly from the set of M attributes. N_{att} is held constant during the rule sets learning.

Every time a new training example is available the algorithm generates a random number i between 0 and 1 for each rule set. If the generated number i is greater than user defined threshold T_{rnd} (the random threshold), the algorithm proceeds with checking whether for each rule set (RS_j) the example is covered by any rule from RS_j , that is if all the literals of the rule are true for the example. The rule is expanded using only the attributes of the RS that the rule belongs to. To expand the rule we use the algorithm proposed in [3]. If a labeled example is not covered by any rule of any RS , the default rule is updated. The expansion of default rule is performed as in the algorithm proposed in [3].

In the experimental study, we consider two classification strategies: Random Rules without Naive Bayes (NB) (denoted as RR) and Random Rules with NB (denoted as RRNB). The RR strategy classifies a test example in the class that maximizes $p(c_k)$. The RRNB strategy classifies a test example with the class that maximizes the posteriori probability given by naive Bayes.

4. EXPERIMENTAL EVALUATION

The main goal of this experimental evaluation is to study the behavior of the proposed algorithm in terms of performance compare the advancement to the previous version.

The evaluation datasets include both artificial and real data. We analyze the error-rate and Bias-variance decomposition.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM SAC '13 March 18-22, 2013, Coimbra, Portugal
Copyright 2013 ACM 978-1-4503-1656-9/13/03 ...\$15.00.

Algorithm 1: Random Rules Algorithm

Input:

M : Number of total attributes;
 N_{rs} : Number of rules sets;
 N_{att} : Number of attributes of each rules set;
 T_{rnd} Random threshold;
 N_{min} Minimum number of examples;

Result: S_{rs} New main set of rules ($S_{rs}.size = N_{rs}$)

begin

```
Let each  $RS \in S_{rs}$  is empty,  $S_{rs} \leftarrow \{\}$ ;  
Let  $defaultRule \zeta \leftarrow 0$ ;  
foreach  $RS \in S_{rs}$  do  
  while number of attributes of  $RS_j < N_{att}$  do  
    Attribute  $\bar{a} = rand(N_{att})$ ;  
     $N_{att} \leftarrow N_{att} \cup \bar{a}$ ;  
  
  foreach example  $(\vec{x}, y_k)$  do  
    foreach  $RS \in S_{rs}$  do  
      /* Generates a random number  
      between 0 and 1 */  
      if  $rand(i) > T_{rnd}$  then */  
        foreach Rule  $r \in RS_j$  do  
          if  $r$  covers the example then  
            Update sufficient statistics of  $r$ ;  
            if Number of examples in  
             $\zeta_{r \bmod N_{min}} = 0$  then  
               $r \leftarrow ExpandRule(r)$ ;  
  
    if No rule from any  $RS \in S_{rs}$  triggers then  
      Update sufficient statistics of the  $defaultRule$ ;  
      if Number of examples in  $\zeta_{\bmod N_{min}} = 0$  then  
        foreach  $RS \in S_{rs}$  do  
           $RS \leftarrow RS \cup ExpandRule(defaultRule)$ ;
```

4.1 Experimental Results

We compare the performance of our algorithms with two other algorithms existing in the literature, VFDR and VFDR_{NB} [3] (in this paper denoted as RC and RCNB respectively).

The Figure 1 summarize the error-rates of classifiers which support our hypothesis that the randomization of the process is beneficial for the performance of the rule classifiers. The analysis of Bias-variance is reported in the Figure 2. This analysis shows that RR and RRNB generate stable predictive models.

We conducted experiments in real data. After the analysis we conclude that the classifier RR and RRNB improve the performance of the VFDR classifiers.

5. CONCLUSION

Learning random rules from data stream is an interesting approach that has not been widely explored by stream mining community. In this paper we introduced a new random rules classification approach for streaming data (RR), which incrementally learns from incoming examples and expands rules over time. RR constructs several sets of rules. The variability between the rules of each set of rules is assured by associating each set of rules to a set of N_{att} attributes randomly generated from the full attribute set. In order to determine the class of an instance, the method combines the results of several sets of rules through a voting mechanism. The experimental results in artificial and real datasets clearly show the

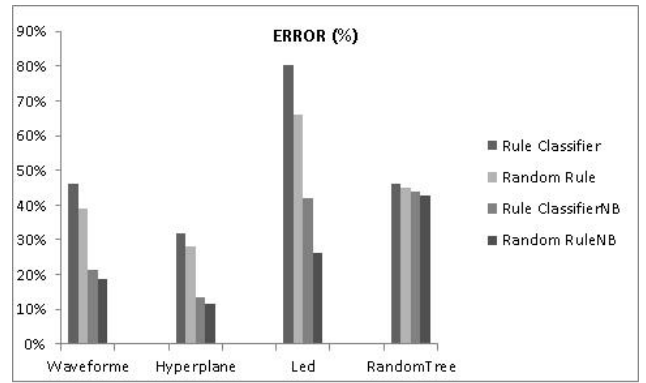


Figure 1: Error rate of classifiers

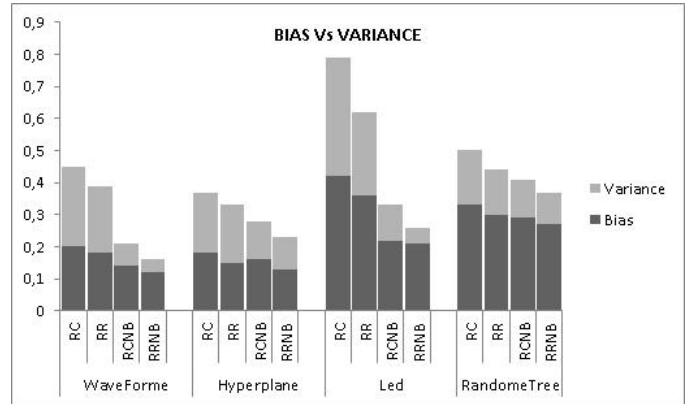


Figure 2: Bias and variance of classifiers

advantages of this approach. RR is more competitive in terms of error-rate and in terms of the bias-variance analysis than RC, the state-of-the-art in streaming decision rule learning. RR generates very stable predictive models concerning variations of the training set. We are working in extending this work to time-changing streams.

Acknowledgments: The authors acknowledge the financial support given by the project FCT-KDUS(PTDC/EIA/098355/2008), the ERDF through the COMPETE Programme and by National Funds through FCT within the project FCOMP - 01-0124-FEDER-022701.

6. REFERENCES

- [1] L. Breiman. Random forests. *Machine Learnin*, 45(1):5–32, 2001.
- [2] J. Cendrowska. Prism: an algorithm for inducing modular rules. *International Journal of Man-Machine Studies*, pages 27(4): pp. 349–370, 1987.
- [3] J. Gama and P. Kosina. Learning decision rules from data streams. In *IJCAI*, pages 1255–1260. AAAI, Menlo Park, USA, 2011.
- [4] M.A.Bramer. An information-theoretic approach to the pre-pruning of classification rules. In *Intelligent Information Processing*, pages pp. 201–212, Kluwer, 2002.
- [5] F. Stahl and M. Bramer. Random prism: An alternative to random forests. In *ICITAAI*, pages pp. 5–18. Cambridge, UK, 2011.