Evolutionary Pruning of Non-Nested Generalized Exemplars

Daniela Zaharie, Lavinia Perian, Viorel Negru and Flavia Zamfirache
Department of Computer Science
West University of Timişoara, blvd. Vasile Pârvan, 300223 Timişoara, Romania
Email: {dzaharie,perian,vnegru,zflavia}@info.uvt.ro

***Abstract*—This paper investigates the ability of an evolutionary pruning mechanism to improve the predictive accuracy of a classifier based on non-nested generalized exemplars. Two pruning algorithms are proposed: one which selects the most representative generalized exemplars and the other one which simultaneously selects both relevant exemplars and relevant attributes. Experimental studies conducted for a set of twenty-one datasets illustrated that both algorithms induce a significant improvement on the classification ability of the selected set of non-nested generalized exemplars.**

I. Introduction

Inducing small and accurate classification models from data is an important issue in machine learning. Finding the right balance between the predictive abilities of the model and its size is a difficult problem. However, sometimes reducing the model size implicitely leads to a more effective classifier. The model size can be controlled either in a pre-processing step (e.g. instance selection, attribute selection) or in a post-processing step (e.g. decision trees pruning, rules sets pruning).

In this paper we address the problem of pruning the set of generalized exemplars induced using a hybrid instance based learner, i.e. NNGE (Non-Nested Generalized Exemplars [7]). The pruning process is based on an evolutionary approach inspired by a technique recently proposed in [5] and which, despite its simplicity, proved to be very effective in improving the classification ability of a simple heuristic extractor of generalized exemplars from the training set. The natural question which arises is if this evolutionary selection can be used to improve the effectiveness of other classifiers based on generalized exemplars. Therefore, in this paper we investigate the possibility of improving the behavior of NNGE by evolutionary selection of generalized exemplars and attributes.

The rest of the paper is organized as follows. Section 2 presents the particularities of the classification based on generalized exemplars with an emphasis on NNGE and describes the changes we propose on NNGE splitting mechanism. In Section 3 is presented the problem of pruning the classification models and the evolutionary selection proposed in [5]. The proposed algorithms are presented in Section 4 while the results of an experimental study on 21 datasets are presented and discussed in Section 5. Section 6 concludes the paper.

II. Classification Based on Generalized Exemplars

The classifiers based on generalized exemplars are hybrid instance based learners which combine the idea of nearest neighbours classifiers and that of rule based classifiers. The element borrowed from the nearest neighbours methods is to use the distance to prototypes when deciding to which class a given instance belongs. Unlike the case of pure instance based learners where the prototypes coincide with the training instances, in the case of classifiers based on generalized exemplars the prototypes are sets of instances which can be interpreted as rules. On the other hand the matching between an instance and a generalized exemplar should not be necessarily exact but it could be partial, being based on the computation of a specific distance.

One of the first methods implementing the idea of the hybrid instance based learning is the Nested Generalized Exemplar (NGE) theory [9] which uses both simple instances and generalized exemplars represented as axes-parallel hyperrectangles to model the concepts. In the NGE learning the examples (training instances) are incrementally generalized leading finally to a set of generalized exemplars and a possible set of non-generalized exemplars (hyperrectangles consisting of a single training instance). In the initial versions of NGE the generalized exemplars can overlap and can be nested (corresponding to rules with exceptions). Further investigations led to the conclusion that avoiding the overlapping [10] and/or the nesting [7] can improve the classification performance of classifiers based on generalized exemplars. Currently the versions which are still used in practice are BNGE (Batch Non-overlapping Generalized Exemplars) proposed in [10] and NNGE (Non-Nested Generalized Exemplars) proposed in [7] and implemented in the Weka toolkit [11]. In this paper we focus on NNGE and investigate the possibility of improving its classification performance by evolutionary selection of relevant hyperrectangles and attributes.

A. The General Structure of NNGE

In order to illustrate the NNGE learning process let us consider a set of L training instances (examples), $(E^1, E^2, \ldots, E^L)$, each one containing the values of N attributes (numerical or nominal). The aim of the learning process is to construct a set of generalized exemplars (hyperrectangles), $\mathcal{H} = \{H^1, H^2, \ldots, H^K\}$. A hyperrectangle

978-1-4244-9107-0/11/$26.00 ©2011 IEEE

usually covers a set of training instances belonging to the same class and each of its dimensions is specified either by a range of values (in the case of numerical attributes) or by an enumeration of values (in the case of nominal attributes). If during the learning process a hyperrectangle corresponding to a given class covers a training instance belonging to a different class then this training instance is considered to be a conflicting example. The learning process is incremental, for each example E^j the following three main steps being applied: *classification* (the hyperrectangle H^k which is closest to E^j is identified by using a distance-based criterion), *model adjustment* (the hyperrectangle H^k is split if it covers a conflicting example) and *generalization* (if it is possible, H^k is extended in order to cover E^j). The general structure of the NNGE is described in Algorithm 1.

Algorithm 1 The general structure of NNGE

$\mathcal{H} \leftarrow \emptyset$
for $j \in \{1, \ldots, L\}$ **do**
  if $\mathcal{H} = \emptyset$ **then**
    $\mathcal{H} \leftarrow \mathcal{H} \cup E^j$
  else
    Find $H^k \in \mathcal{H}$ such that $D(H^k, E^j) \leq D(H^q, E^j)$, for all $H^q \in \mathcal{H}$
    if $D(H^k, E^j) = 0$ **then**
      if $class(H^k) \neq class(E^j)$ **then**
        $H_k \leftarrow$ Split(H^k, E^j)
      end if
    else
      $H' \leftarrow$ Extend(H^k, E^j)
      if H' overlaps with conflicting hyperrectangles **then**
        $\mathcal{H} \leftarrow \mathcal{H} \cup \{E^j\}$
      else
        $\mathcal{H} \leftarrow \mathcal{H} \backslash \{H^k\} \cup \{H'\}$
      end if
    end if
  end if
end for

The classification step is based on the computation of the distance $D(E, H)$ between an example $E = (E_1, E_2, \ldots, E_N)$ and a hyperrectangle H defined by its components $(H_1, H_2, \ldots, H_N)$ as given in Eq. (1). The component H_i of a hyperrectangle is an interval $[H_i^{min}, H_i^{max}]$ in the case of a numerical attribute and a finite set of values in the case of a nominal attribute.

$$D(E, H) = \sqrt{\sum_{i=1}^{N} w_i \frac{d(E_i, H_i)}{E_i^{max} - E_i^{min}}} \quad (1)$$

The distances between the examples attributes and the hyperrectangles "sides" are given by Eq. (2) for numerical attributes and by Eq. (3) for nominal attributes, respectively. The parameters w_i are weights corresponding to attributes and are computed based on the mutual information between the attribute and the class.

$$d_{num}(E_i, H_i) = \begin{cases} 0 & \text{if } E_i \in [H_i^{min}, H_i^{max}] \\ E_i - H_i^{max} & \text{if } E_i > H_i^{max} \\ H_i^{min} - E_i & \text{if } E_i < H_i^{min} \end{cases} \quad (2)$$

$$d_{nom}(E_i, H_i) = \begin{cases} 0 & \text{if } E_i \in H_i \\ 1 & \text{if } E_i \notin H_i \end{cases} \quad (3)$$

B. Avoiding the Exemplars Overlapping and Nesting

In order to avoid the existence of overlaped hyperrectangles having different classes, the generalization (extension) of a hyperrectangle is accepted only if the new hyperrectangle does not overlap with hyperrectangles having a different class. If there is an overlap, the generalization process is abandoned and the training instance is added to the model as a non-generalized exemplar.

On the other hand, in order to avoid the generation of nested hyperrectangles, NNGE adjusts the hyperrectangle H which covers a conflicting example E, i.e. $D(E, H) = 0$ and $class(H) \neq class(E)$, such that this example is no more covered. This is realized by splitting the hyperrectangle in a few other hyperrectangles and potentially some isolated instances. This is one of the critical components of NNGE and makes NNGE different with respect to other methods based on generalized exemplars.

The splitting process consists of changing one of the dimensions of the hyperrectangle. Since one of the goals is to limit as much as possible the number of exemplars (especially of non generalized ones) the choice of splitting attribute should take this into account. In the case of nominal attributes this is ensured by choosing the attribute for which the value in the conflicting example is less frequent amongst the other examples included in the hyperrectangle.

In the case of numerical attributes several criteria to choose the splitting attribute can be identified. In our implementation we used four criteria: (i) "closest margin"; (ii) "balanced split"; (iii) "maximal bandwidth"; (iv) "minimal bandwidth". The first one ("closest margin") corresponds to the variant of NNGE implemented in Weka and consists of choosing the attribute for which the corresponding value of the conflicting attribute is the closest to a margin of the covering hyperrectangle (Fig. 2). In the case of a tie, the attribute leading to the largest number of training examples included in one of the splitting hyperrectangles is chosen. The second variant ("balanced split") chooses the attribute for which the ratio of the distances between the value of the conflicting example and the hyperrectangle margins is as close as possible to 1 (Fig. 1). This approach would lead to a more balanced split than the previous one. The last two variants are based on analyzing the size of the "free" space between the resulting hyperrectangles (this space would contain only the examples having exactly the same value as the conflicting instance for the splitting attribute). Fig. 2 illustrates the case when the

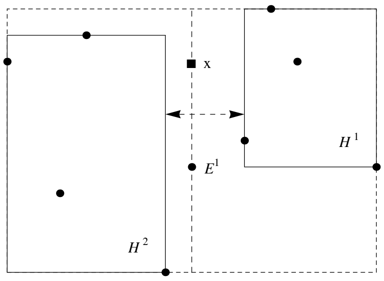

Fig. 1. Splitting the initial hyperrectangle (dashed border) by the first attribute (the conflicting instance is denoted by x)

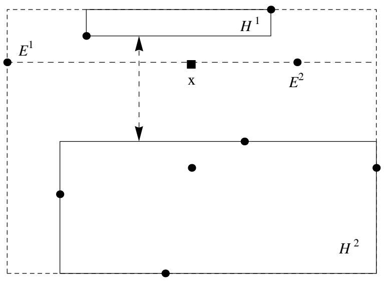

Fig. 2. Splitting the initial hyperrectangle (dashed border) by the second attribute (the conflicting instance is denoted by x)

maximal bandwidth is chosen (corresponding to the second attribute) while Fig. 1 illustrates the case of the minimal bandwidth (corresponding to the first attribute). In all four cases the initial hyperrectangle is divided in at least two hyperrectangles: one for the examples having the value of the splitting attribute strictly higher than the value of the conflicting instance (H^1 in Figs. 1,2) and one containing the examples corresponding to strictly smaller values (H^2 in Figs. 1,2). The examples having the same value of the splitting attribute as the conflicting instance will either form a different hypperrectangle or will remain as non-generalized exemplars. Since there is no best splitting for all cases we used a combined version: at each splitting stage all four variants are tried and that leading to the smallest number of non-generalized examplars is chosen. A preliminary analysis (not presented in this paper) illustrated that this approach improves the behavior of the original NNGE [7] which uses only the "closest margin" variant. In the case of mixed attributes the splitting one is chosen between the best nominal and the best numerical attributes, based on the same criterion: minimization of the number of non-generalized exemplars.

III. Pruning of Classification Models and Related Work

There are two main approaches in reducing the size of classifiers: pre-pruning and post-pruning. The main examples of pre-pruning approaches are those aiming to select (prune, edit) the "good" and "clean" training instances (or prototypes) and those aiming to select the relevant attributes. Post-pruning is applied to the induced model, e.g. decision trees pruning or rules pruning.

A. Instance and Attribute Selection

The number of existing techniques for instance and/or attributes selection is impressing. Since the selection is a hard combinatorial optimization problem the evolutionary algorithms proved to be effective. For instance, evolutionary attribute selection is able to deal well with the interaction between attributes [4] which explains the large number of evolutionary attribute selectors. A similar situation arises in the case of instance selection (for a recent review on evolutionary instance selection see [2]). The evolutionary algorithms proved also to be effective for simultaneous selection of instances and attributes [8].

B. Evolutionary Selection of Nested Generalized Exemplars

In [5] was proposed a first approach for evolutionary selection of hyperrectangles, called EHS-CHC. The selection process starts from a set of hyperrectangles generated using a simple heuristic: for each instance in the training set the hyperrectangle covering the closest k nearest neighbours belonging to the same class (the $k+1$ closest nearest neighbor should belong to a different class) as the processed instance is constructed. The generated hyperrectangles can overlap or can be nested. The selection process is based on the evolution of a population of binary encoded elements corresponding to various subsets of the initial set of hyperrectangles. Despite the simplicity of the approach it proved to be surprinsingly effective both with respect to the predictive accuracy of the resulting classifier and with respect to the reduction of the number of hyperrectangles. Recently, an improved variant, based on a preliminary noise filtering of the training set, has been proposed and analyzed for a large number of datasets [6].

IV. Evolutionary Pruning in NNGE

Once a set $\mathcal{H} = \{H^1, H^2, \ldots, H^K\}$ of hyperrectangles has been generated by the NNGE algorithm it can be post-processed in order to reduce its size and, hopefully, to improve the classification accuracy. Following the idea of the hyperrectangles selection presented in [5] and extended in [6] we developed an evolutionary pruning algorithm acting as post-processor of the results produced by NNGE. The first version of the algorithm, called EP-NNGE (Evolutionary Pruning in NNGE) is based on the idea of evolving a population of M binary strings containing K components. Each element, x, of the population corresponds to a subset of $\mathcal{H}$, e.g. if a component x_k has the value 1 it means that H^k is selected into the model, while if it is 0 it means that H^k is not selected. The quality of an element x is quantified, as in EHS-CHC, using two measures: one related to the accuracy of the classifier based on the selected hyperrectangles ($\mathcal{H}(x)$) and the other one related to the reduction of the model size. Thus the fitness is given by Eq. (4) where Acc denotes the accuracy, $|\mathcal{H}|$ denotes the number of hyperrectangles and $\lambda \in (0,1)$ is a parameter controlling the compromise between the two quality measures.

$$f(x) = \lambda Acc(\mathcal{H}(x)) + (1-\lambda)\frac{|\mathcal{H}| - |\mathcal{H}(x)|}{|\mathcal{H}|} \quad (4)$$

The evolutionary algorithm is inspired by the CHC adaptive search algorithm [3], also used in [5] and [6]. The general structure of the algorithm is described in Algorithm 2.

Algorithm 2 The general structure of EP-NNGE

Random initialization of the population $X(0) \leftarrow \{x_1(0), \ldots, x_M(0)\}$
$g \leftarrow 0$
Evaluate $X(0)$
while the stopping condition is false **do**
  Copy $X(g)$ to $X(g+1)$
  for $i \in \{1, \ldots M\}$ **do**
    Select two random parents $x_{j_1}(g)$ and $x_{j_2}(g)$
    if Hamming$(x_{j_1}(g), x_{j_2}(g)) > K/4$ **then**
      Construct $y_{j_1}(g)$ and $y_{j_2}(g)$ from the parents by uniform crossover
      Evaluate $y_{j_1}(g)$ and $y_{j_2}(g)$ and append them to $X(g+1)$
    end if
  end for
  Truncate $X(g+1)$ by retaining its best M elements
  $g \leftarrow g+1$
end while

The second approach we propose in this paper is that of simultaneously selecting hyperrectangles and attributes. In this case each element in the population has $K+N$ components (K being the initial number of hyperrectangles and N being the total number of attributes). The corresponding algorithm (EPA-NNGE) has the same structure as EP-NNGE and the population elements are evaluated also by using Eq. (4). The main difference between EPA-NNGE and EP-NNGE is related to the computation of the classification accuracy: in the computation of the distance between a test instance and a hyperrectangle, all non-selected attributes (as are they specified by the corresponding part x^a of the population elements) are just ignored, meaning that instead of Eq. (1) is used the distance given by Eq. (5).

$$D_x(E,H) = \sqrt{\sum_{i=1}^{N} x_i^a w_i \frac{d(E_i,H_i)}{E_i^{max} - E_i^{min}}} \qquad (5)$$

V. Experimental Results

The aim of the experimental analysis was to assess the ability of the evolutionary pruning algorithms (EP-NNGE and EPA-NNGE) to improve the classification accuracy of NNGE and to reduce the size of the induced model (the number of hyperrectangles).

A. The Experimental Setup

The experiments were conducted using 21 datasets from KEEL repository [1]. These datasets are constructed based on those at UCI Machine Learning Repository (http://www.ics.uci.edu/ mlearn/ MLRepository.html) but they are already sliced in folds (e.g. 10) in order to allow a cross-validation based comparison between different methods. These datasets were selected in order to allow us to compare the results we obtained with those reported in [6]. The characteristics of the datasets (number of instances, attributes and classes) are presented in the first columns of Table I. Both EP-NNGE and EPA-NNGE were used with $M = 50$ elements in the population and as stopping criterion was used a maximal number of generations (100) combined with a maximal number of generations without progress (50). These values are consistent with those used in [6] ($M = 50$, 200 generations). No parameter tuning and no effort to improve the behavior of the evolutionary algorithm was done in order to illustrate that a simple evolutionary selection which does not involve many computational resources can improve significantly the behavior of NNGE. All estimations of the classification accuracy and of the model size were obtained by 10-fold cross-validation, and for each training/testing pair the evolutionary selection process was independently applied for 10 times. Based on all 100 values the average and the standard deviation values were computed.

B. Influence of Evolutionary Pruning on the NNGE Classification Performance

In order to assess the positive impact of the evolutionary pruning on the classification performance of NNGE the average accuracy on the testing datasets obtained by the proposed algorithms (EP-NNGE and EPA-NNGE) was compared with the accuracy obtaing by the non-pruned version of NNGE and with the best corresponding results reported in [6]. The results are presented in Table I. In all cases, EP-NNGE and EPA-NNGE led to significantly higher accuracy than the non-pruned NNGE. All averaged accuracies obtained by EPA-NNGE are higher than the best values reported in [6] and in 17 out of 21 the superiority proved to be statistically significant. The statistical analysis was based on a Student test with 0.01 as level of significance. The significantly better values are boldfaced in Table I (when two values on the same row are boldfaced it means that the difference between them is not significant). Concerning the comparison between EP-NNGE and EPA-NNGE, they proved to behave similarly in 15 out of 21 cases. In 5 cases EPA-NNGE led to significantly better results than EP-NNGE while in only one case EP-NNGE was significantly better than EPA-NNGE.

Since in this experiment the focus was on improving the classification accuracy, the evolutionary process was guided by a fitness function based on a high value for λ (e.g. $\lambda = 0.995$). The influence of the parameter λ on the accuracy and on the reduction of the model size (the hyperrectangle reduction ratio is defined as $(|\mathcal{H}| - |\mathcal{H}(x_{best})|)/|\mathcal{H}|$) has been analyzed for several datasets obtaining similar patterns of behavior. Results corresponding to two data sets ("breast" and "haberman" are illustrated in Figs. 3 and 4). As it would be expected, by increasing the value of λ the value of accuracy also increases while the reduction ratio decreases. An interesting aspect is the fact that for all analyzed cases the accuracy corresponding to the non-pruned NNGE is obtained for λ near 0.2, while for

TABLE I
CLASSIFICATION ACCURACY (%)

No.	Dataset	#Inst.	#Num.	#Nom.	#Cl.	NNGE (avg±stdev)	EP-NNGE (avg±stdev)	EPA-NNGE (avg±stdev)	Best rule based classifier [6] Acc.	Method
1	appendicitis	106	7	0	2	80.36±7.6	**93.94**±6.8	**96.89**±6.3	86.91±11.5	Filtered EHS-CHC
2	australian	690	8	6	2	85.21±5.5	**92.20**±3.0	**94.55**±2.7	85.80±2.9	Filtered EHS-CHC
3	breast	286	0	9	2	73.75±7.2	**90.44**±6.4	**94.10**±5.7	73.80±6.2	EHS-CHC
4	bupa	345	6	0	2	59.94±6.7	**87.88**±3.4	**90.28**±5.2	65.47±4.5	BNGE
5	cleveland	297	13	0	5	52.31±7.9	**77.97**±7.0	**82.93**±6.6	56.81±6.5	EHS-CHC
6	contraceptive	1473	9	0	3	44.94±3.6	**75.40**±4.3	70.80±4.9	53.23±5.2	RIPPER
7	crx	125	6	9	2	81.96±5.6	**93.85**±3.5	**94.33**±10.4	84.35±4.7	C4.5Rules
8	dermatology	366	34	0	6	96.90±3.3	97.73±2.9	**99.34**±1.2	97.00±2.4	Filtered EHS-CHC
9	ecoli	336	7	0	8	83.32±4.2	**89.87**±4.9	**90.54**±3.9	82.16±4.6	BNGE
10	glass	214	9	0	7	68.70±13.5	76.09±8.9	**86.42**±9.0	73.61±11.9	1NN
11	haberman	306	3	0	2	68.32±7.0	**93.38**±4.4	**90.85**±6.4	74.49±6.0	Filtered EHS-CHC
12	hepatitis	155	19	0	2	86.00±11.3	**93.07**±7.6	**98.35**±4.5	83.88±6.9	RISE
13	iris	150	4	0	3	96.00±4.7	98.00±3.2	99.33±2.1	96.67±4.7	Filtered EHS-CHC
14	led7digit	500	7	0	10	63.40±6.8	72.52±6.1	**78.28**±6.2	71.40±4.8	C4.5Rules
15	mammographic	961	5	0	2	73.66±4.8	**92.29**±3.6	**89.73**±6.2	83.04±4.4	INNER
16	newthyroid	215	5	0	3	93.07±3.9	96.91±3.3	97.79±2.3	97.23±2.3	1NN
17	pima	768	8	0	2	73.96±3.2	**88.94**±4.1	**89.69**±3.9	75.01±3.6	EHS-CHC
18	sonar	208	60	0	2	59.11±9.7	81.99±9.7	**94.35**±6.2	85.55±7.5	1NN
19	wine	178	13	0	3	92.15±4.6	95.52±3.5	**99.44**±1.8	96.60±2.9	BNGE
20	wisconsin	683	9	0	2	96.79±2.1	98.78±1.5	99.10±1.1	97.00±2.8	BNGE
21	zoo	101	16	0	7	95.16±6.6	96.00±4.9	100±0	96.83±5.2	RISE

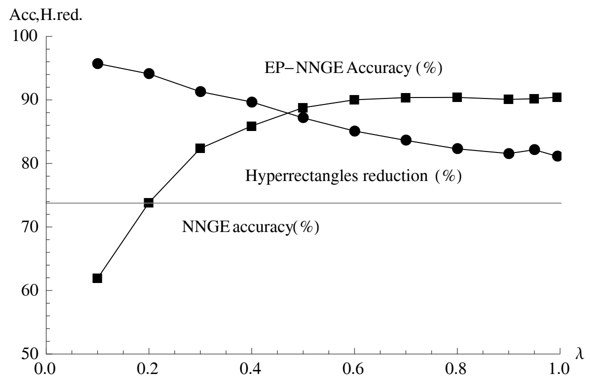

Fig. 3. Influence of λ on the EP-NNGE accuracy gain and hyperrectangles reduction. Dataset: breast

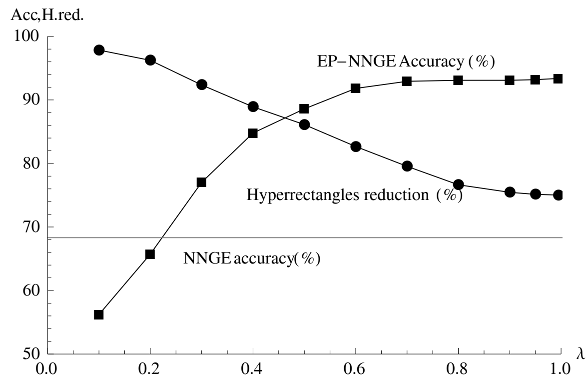

Fig. 4. Influence of λ on the EP-NNGE accuracy gain and hyperrectangles reduction. Dataset: haberman

λ close to 0.5 the accuracy and the model reduction ratio are almost equal.

C. *Influence of Evolutionary Pruning on the Model Reduction*

Even if the focus of the experiments was on improving the classification performance we also remarked that both EP-NNGE and EPA-NNGE led to a significantly smaller number of hyperrectangles than the non-pruned NNGE. On the other hand, when we compare EP-NNGE with EHS-CHC [5] and Filtered EHS [6] the difference is not so significant. For instance the number of hyperrectangles induced by EP-NNGE is smaller than that induced by EHS-CHC (based on $\lambda = 0.5$) for 16 datasets, but is significantly higher in the case of three datasets (”contraceptive”, ”led7digit”, ”mammographic”). This result can be explained by the fact that these three datasets contain a large number of instances. However, when compared to filtered-EHS, a statistical analysis based on the Wilcoxon signed rank test revealed that there is no overall significant difference between the number of hyperrectangles selected by EP-NNGE and by the filtered-EHS. The influence of the evolutionary pruning implemented in EPA-NNGE on the induced model performance and size is graphically illustrated for all 21 datasets in Figs. 5 and 6. Figure 5 illustrates the accuracy gain (computed as $(Acc(\text{EPA-NNGE}) -$

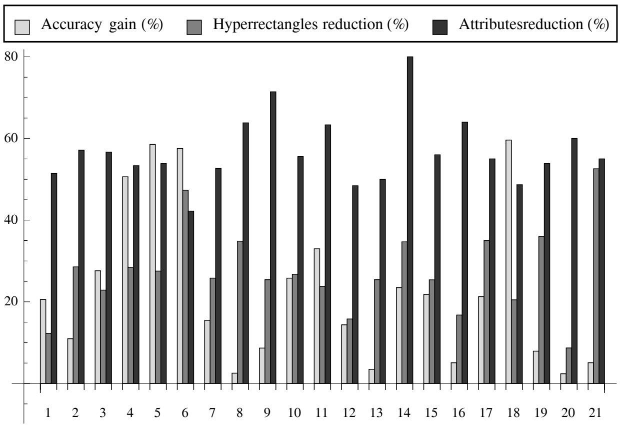

Fig. 5. EPA-NNGE vs. NNGE: accuracy gain(%), hyperrectangles reduction(%) and attributes reduction(%) for 21 datasets

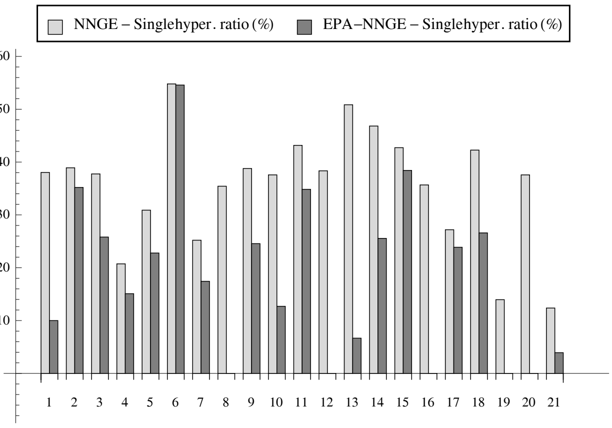

Fig. 6. EPA-NNGE vs. NNGE: ratio (%) of non generalized exemplars (single point hyperrectangles) for 21 datasets

$Acc(\text{NNGE}))/Acc(\text{NNGE}) \cdot 100$), the ratio of selected hyperrectangles (computed as $|\mathcal{H}_{\text{EPA-NNGE}}|/|\mathcal{H}_{\text{NNGE}}| \cdot 100$) and the ratio of selected attributes (computed as $N_{\text{EPA-NNGE}}/N \cdot 100$). The largest gain in accuracy (around 50%) was obtained for "bupa", "cleveland", "contraceptive" and "sonar" datasets. This can be explained by the fact that NNGE provided low accuracy for these data. On the other hand the smallest gain in accuracy (smaller than 10%) was obtained for "dermatology", "iris", "newthyroid", "wisconsin" and "zoo" datasets (for all of these, NNGE led to a rather high classification accuracy, leaving no room for significant improvements). The number of hyperrectangles was reduced with at least 50% for all datasets. The highest reduction (the number of selected hyperrectangles is less than 20% of the initial number) was obtained for "appendicitis", "hepatitis", "newthyroid" and "wisconsin" datasets. On the other hand the smallest reduction (50%) was obtained for "contraceptive" and "zoo". The number of selected attributes is between 40% and 60% of the initial number for almost all datasets. The smallest reduction in the number of attributes was remarked for "ecoli" and "led7digit" datasets. Another analysis we conducted was related to the importance of non-generalized hyperrectangles. The ratio between the number of non-generalized exemplars and the total number of exemplars is illustrated in Figure 6 both for NNGE and EPA-NNGE. For all datasets the ratio of non generalized exemplars is smaller in the case of EPA-NNGE than in the case of NNGE. There are also five datasets ("dermatology", "hepatitis", "newthyroid", "wine" and "wisconsin") for which all non generalized exemplars were eliminated by the evolutionary pruning process. This suggests that for these datasets the non generalized exemplars do not play a significant role.

VI. Conclusions and Further Work

As experimental results suggest the evolutionary pruning of hyperrectangles induced by NNGE leads to a significant improvement of the classification performance and to a reasonable reduction of the model size. The simultaneous selection of hyperrectangles and attributes proved to be very effective, i.e. the classification accuracy is higher than that obtained by the best rule-based classifiers identified in the experimental analysis conducted in [6]. All these results were obtained by applying a simple evolutionary algorithm and a simple aggregation approach in dealing with the multi-objective character of the problem (both the accuracy and the model size should be optimized). This means that there is still room for improvements by changing the evolutionary approach. Further work will also address the scalability issue in the context of the pruned NNGE.

Acknowledgment

This work was supported by the grants POSDRU 107/1.5/G/13798 and PNCDII 12-122/01.10.2008(ASISTSYS).

References

[1] J. Alcal-Fdez et al, *KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework*, Journal of Multiple-Valued Logic and Soft Computing, Vol. 17:2-3, pp.255-287, 2011.

[2] J. Derrac, S. Garcia, F. Herrera, *A Survey on Evolutionary Instance Selection and Generation*, International Journal of Applied Metaheuristic Computing, 1(1), pp. 60-92, 2010.

[3] L.J. Eshelman, *The CHC adaptive search algorithm: How to safe search when engaging in nontraditional genetic recombination*, in G.J.E. Rawlings, FOGA, pp. 265-283, 1991.

[4] A. Freitas, *A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery*, in Advances in Evolutionary Computing, Springer-Verlag NY, 2003.

[5] S Garcia et al, *A First Approach to Nearest Hyperrectangle Selection by Evolutionary Algorithms*, Proc. of 9th Intern. Conf. on Intelligent System Design and Applications, Pisa, Italy, pp.517-522, 2009.

[6] S. Garcia et al, *Evolutionary Selection of Hyperrectangles in Nested Generalized Exemplar Learning*, Applied Soft Computing, Vol.11(3), pp. 3032-3045, 2011.

[7] B. Martin, *Instance-Based Learning: Nearest Neighbour with Generalisation*, Working Paper Series 95/18 Computer Science, Hamilton,University of Waikato, pp. 90, 1995.

[8] F. Ros, S. Guillaume, M. Pintore, J.R. Chrétien, *Hybrid genetic algorithm for dual selection*, Pattern Analysis Applications, Vol.11, pp. 179-198, 2008.

[9] S. Salzberg, *A Nearest Hyperrectangle Learning Method*, Machine Learning, Vol.6, pp. 251-276, 1995.

[10] D. Wettschereck and T.G. Dietterich, *An Experimental Comparison of the Nearest-Neighbor and Nearest-Hyperrecyangle Algorithms*, Machine Learning, Vol. 19, pp. 1- 25, 1995.

[11] I.H. Witten and E. Frank, *Data Mining. Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, 2005.