Reduced Reward-punishment editing for building ensembles of classifiers

Loris Nanni, Annalisa Franco

Expert Systems and Applications (JCR: 1.924)

February 1, 2012

Paper outline

1. Introduction
   1. Constructing ensembles: pattern perturbation, feature perturbation, classifier perturbation, hybrid methods
   2. Editing techniques: WPE and Reward-punishment
2. Proposed system
   1. Reduced Reward-punishment editing
   2. Variants to known ensemble methods: bagging-variant, RF-variant and Input-Decimated Ensemble-variant
3. Experimental results
   1. Settings
   2. Experiment 1
   3. Experiment 2
4. Conclusions

Classic Approaches for constructing Multi-Classifiers

- Pattern perturbation: instead of the original training set, modified versions of it are fed to the different classifiers. Examples: bagging, arcing, boosting, ...
- Feature perturbation: new training sets are built with different feature sets. Examples: random subspace, cluster-based pattern discrimination, input decimated ensemble, ...
- Classifier perturbation: same training set, but classifiers in the ensemble have different parameters or belong to a different class of classifiers.
- Hybrid methods: more than one class of perturbation is used. Examples: random forests, rotation forests, rotboost,...

Editing techniques

- Initial motivation: alleviate the CPU and memory requirements of k-NN classifiers.
- Goal: remove noisy samples from the training set using some metric
- Examples:
  - WPE (Paredes & Wagner, 2000):
    - samples are weighted using the ratio of (dissimilarity between the sample and others of the same class) and (dissimilarity between the sample and others of different class)
    - samples with higher weights are discarded
  - Reward-punishment editing (in press):
    - global criterion: rewards samples correctly classified by a k-NN rule (on a set of prototypes)
    - local criterion: rewards samples that contribute to correctly classify their neighbours
    - punishes other samples, has a lot of parameters

Reduced Reward-punishment editing (RPP)

- RPP
  - only uses local criterions and two parameters: α and *et*.
- Samples x_i are given two weights:
  - WR(i): number of times x_i has contributed to the correct classification of another pattern
  - WP(i): number of times x_i has contributed to a wrong classification of another pattern
- These weights are normalized
- Final weight:

$$\boldsymbol{WF}(i) = \alpha \times \boldsymbol{WR}(i) + (1 - \alpha) \times (1 - \boldsymbol{WP}(i))$$

- *et* percent of patterns wight highest WF weights are retained

RPP II

```
RP-EDITING(TS, CL, k, α, et)
begin
    WR = WP = WPR = 0
    for each x_i ∈ TS
        // k-NN classification of the pattern x_i
        [L, c] = K-NN(x_i, TS, k)
        // Is the pattern x_i correctly classified?
        if CL(i) = c then
            // Reward of the patterns that contributed to the correct classification of x_i
            for j = 1 to k
                if CL(L(j)) = c then
                    WR(L(j))= WR(L(j))+ 1
                end if
            end for
        else
            // Punishment of the patterns that contributed to the wrong classification of x_i
            for j = 1 to k
                if CL(L(j)) = c then
                    WP(L(j)) = WP(L(j))+ 1
                end if
            end for
        end if
    end for
    NORMALIZE(WP, WR)
    // Computation of the final weight and Editing
    for each x_i ∈ TS
    WF(i)=α×WR(i)+(1- α)×(1-WP(i))
    RANKANDEDIT(TS, WF, et)
end
```

RRP as a pattern-perturbation technique

- combinations of different values of these parameters can be used to generate different training sets.
  - $\alpha \in [0, 0.25, 0.5, 0.75, 1]$
  - $et \in [10\%, 22.5\%, 35\%, 47.5\%]$
  - k-NN parameter: $k \in [1, 3, 5, 7, 9]$

Ensemble methods

- Bagging-variant (*RP*): original Bagging (Breinman, 1996) but, instead of random bootstrapping, sample subsets are generated giving different values to RPP parameters
- Rotation Forest-variant (*EditedRF*): PCA/ICA + same idea
- Input Decimated Ensemble-variant (*EditedID*): PCA/ICA + same idea

Settings

- Goal: validate the choice of an editing algorithm for building an ensemble of classifiers
- Classifiers: Ensembles with three editing techniques (Bagging, WPE and RRP) vs stand-alone classifiers
    - Adaboost.M1
    - 1-Nearest neighbour
- Adaboost is a bi-class classifier, therefore selected databases contained only 2 classes

Results

Dataset	AdaBoost			
	NO	Bagging	WPE	RP
Breast cancer	4.8	4.3	6.2	**3.3**
Heart disease	16.8	14.9	16.5	**13.8**
Ionosphere	10.0	9.3	11.0	**7.7**
Medullo	23.3	**21.6**	23.3	**21.6**
Pima Indian diabetes	27.0	25.8	28.4	**25.7**
WDBC	3.3	3.1	4.3	**2.3**
Sonar	16.5	17.5	20.9	**16.5**
Average	14.5	13.7	15.8	**12.9**

Adaboost.M1

Dataset	Nearest neighbor			
	NO	Bagging	WPE	RP
Breast cancer	4.5	4.4	4.4	**3.1**
Heart disease	36.0	21.6	**15.6**	18.7
Ionosphere	14.0	14.2	**9.7**	12.5
Iris	5.6	6.1	**5.2**	5.3
Medullo	**25.0**	**25.0**	**25.0**	**25.0**
Wine	5.3	5.1	5.7	**5.0**
Pima Indian diabetes	33.0	30.4	**28.7**	29.1
Wdbc	4.8	4.6	5.0	**4.0**
Sonar	15.8	**15.8**	19.6	16.3
Vehicle	37.5	31.9	31.9	**31.5**
Average	18.1	15.9	15.1	**15.0**

1-NN

Settings

- Goal: more wide comparison
- Ensembles (all tested with 50 classifiers): EditedRF, EditedId, stand-alone decision tree with pruning classifier (DTP), stand-alone SVM, ensemble using Bagging, random subspace ensemble (RS), Rotation Forest with M=3 + PCA/ICA (RF-PCA/RF-ICA), RF-ICA with M=number_of_features/2 (RF-M), IDE-PCA (ID), improved ID where classes are partitioned in clusters (ID-ICA) and RotationBoost + PCA/ICA (RotB-PCA/RotB-ICA)
- Additional statistical test: Wilcoxon Signed-Rank test: EditedRF vs RF-ICA and EditedID vs ID-ICA. Null hypothesis (no difference between accuracies of two ensembles) is rejected.

Results

Method	MEDULLO	WINE	IRIS	IONO	BREAST	HEART	VEIC	DIAB	WDBC	SONAR	AVG
EditeDIF	21.6	2.8	5.6	4.3	3.5	15.3	24.6	24.2	2.4	17.1	12.1
EditeDID	21.6	2.6	4.2	5.8	3.7	17.3	16.7	24.7	3.5	22.5	12.3
DTP	35.0	14.3	7.3	10.9	6.9	24.0	30.3	31.7	6.1	28.9	19.5
SVM	23.3	3.8	3.7	4.9	4.9	17.3	25.8	23.5	4.0	16.6	12.8
Bagging	26.6	13.1	5.5	7.0	4.9	21.3	30.3	24.2	4.6	27.3	16.5
RS	26.6	13.2	6.2	8.1	4.3	18.1	53.0	26.3	4.5	24.7	18.5
RF-Pca	23.3	5.9	6.4	5.8	4.1	15.7	24.9	25.1	2.8	19.0	13.3
RF-Ica	21.6	3.7	4.5	5.4	3.7	15.5	25.2	25.1	2.4	19.8	12.7
RF-M	21.6	2.8	5.6	4.6	3.5	15.4	24.7	24.5	2.4	18.2	12.3
RotB-PCA	21.6	4.6	4.8	4.8	3.6	16.0	21.1	25.9	2.8	18.5	12.4
RotB-ICAa	21.6	3.8	5.8	5.3	3.9	16.8	22.2	25.0	3.1	18.1	12.6
ID	30.0	7.5	4.8	12.9	3.7	24.0	22.2	30.2	5.8	35.0	17.6
ID-PCA	23.3	3.2	4.2	6.0	3.5	16.0	17.4	25.8	3.8	18.5	12.2
ID-Ica	21.6	2.6	5.1	6.9	3.7	17.9	17.3	25.3	3.5	24.0	12.8