Your Paper Title

Author 1, Author 2, Author 3, Author 4

October 31, 2024

Abstract

This is the abstract. Summarize the purpose, methods, and main findings of the study.

1 Introduction

Introduce the problem, background, and significance of the study.

2 Reduction Methods

In this section, we briefly describe reduction techniques employed for instance-based learning. We use a simple 2D dataset for illustrative purposes.

2.1 GCNN

2.2 EENTH

This subsection outlines the Elimination Editing with Nearest-neighbor Threshold (EENTH) method [1]. This approach uses a modified k-NN rule, integrating probability-based decisions for instance elimination. The main steps are outlined below.

1. **Probability-based Classification**: For each instance x, calculate the probability $p_i(x)$ of x belonging to each class i based on its k-nearest neighbors. Probabilities are weighted inversely by the distance to each neighbor and normalized:

$$p_i^j = \frac{|\{x_k \in NN_k(x) : y_k = j\}|}{k} \tag{1}$$

$$P_i(x) = \sum_{j=1}^{k} p_i^j \frac{1}{1 + d(x, x^j)} \tag{2}$$

$$p_i(x) = \frac{P_i(x)}{\sum_{j=1}^{M} P_j(x)} \tag{3}$$

2. **Thresholding**: Define a threshold μ to refine classification, we will denote as $p(x)$ the highest probability. Instances near decision boundaries, where $p(x) < \mu$, are identified as candidates for removal.

3. **Elimination**: If an instance x does not match the class with the highest probability, or if its highest class probability falls below μ, it is removed from the dataset, resulting in an edited set $S \subseteq X$.

The EENTH method thus provides a balance between retaining instances with high classification confidence and discarding uncertain instances near decision boundaries.

2.3 DROP3

In this subsection, we describe the basic concepts of the third method in the Decremental Reduction Optimization Procedure (DROP) family, as presented in Section 3 of Wilson et al. [3]. Although we will not delve into every detail, we describe the main ideas of the algorithm and illustrate them on D_1. See **Figure** 1.

1. **Remove noise**: The first step is to remove noisy instances using Edited Nearest Neighbor (EEN) [2], where any instance misclassified by its k-nearest neighbors is removed. The outcome of applying this technique is shown in **Figure** 2, where noise has been removed. We denote the reduced dataset as $T \subseteq D_1$.

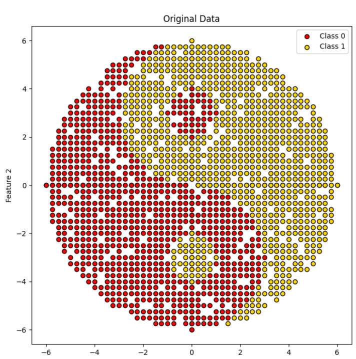

Figure 1: Original Dataset

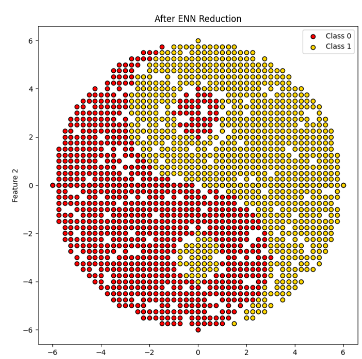

Figure 2: Effect of EEN

2. **Sort points**: The next step is to prioritize removing points that are farthest from the decision boundary. For each point $x_i \in S$ with class y_i, we compute the distance to the nearest point with a different class, denoted as $x_j \in D$ such that $y_j \neq y_i$ and $\nexists x_k : |x_k - x_i| < |x_j - x_i| \wedge y_i \neq y_k$.

3. **Delete points**: Let $S = T$. Starting with the points farthest from the boundary, we check if any associated points (points that have x_i as a neighbor) a_j receive more votes for their correct class with x_i as a neighbor (denoted as with) or if they would be classified correctly if x_i were removed (denoted as without). If without > with, we remove x_j from S, resulting in $S' = S \setminus \{x_j\}$.

4. **Selecting neighbors**: A key distinction between DROP1 and DROP2 is that DROP1 removes points that are removed from the dataset from the list of associates while DROP2 doesn't.

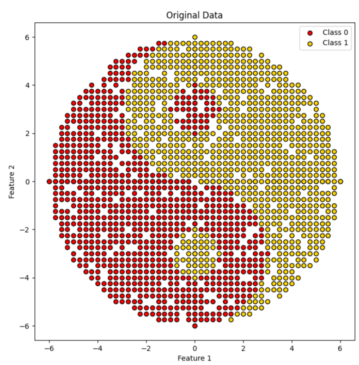

Figure 3: Original Dataset

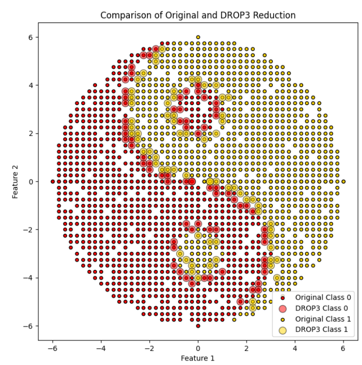

Figure 4: Effect of DROP3

3 Methodology

Describe the datasets and outline each method applied:

3.1 Datasets

Briefly describe Dataset 1 and Dataset 2.

3.2 K-Nearest Neighbors (KNN)

Explain the KNN approach applied.

3.3 Support Vector Machine (SVM)

Describe the SVM approach.

3.4 CGNN Reduction Technique

Outline the CGNN reduction method used.

4 Results

Present the findings for each technique and the statistical analysis:

4.1 KNN Results

Discuss the outcomes for KNN.

4.2 SVM Results

Describe the SVM results.

4.3 CGNN Reduction Results

Present the CGNN findings.

4.4 Statistical Analysis

Detail the statistical analysis performed across each fold.

5 Discussion

Interpret the results, relate to previous work, and discuss implications.

References

[1] Fernando Vázquez, Josep Sánchez, and Filiberto Pla. A stochastic approach to wilson's editing algorithm. pages 35–42, 01 2005.

[2] D. L. Wilson. Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man, and Cybernetics*, 2(3):408–421, 1972.

[3] Dennis R. Wilson and Tony R. Martinez. Reduction techniques for instance-based learning algorithms. *Machine Learning*, 38(3):257–286, 2000.