KNN CLAS

Eduardo Henrique Basilio de Carvalho
Departamento de Engenharia Eletrônica
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
eduardohbc@ufmg.br

Abstract—**This document proposes and analyzes an alternative to the NN-CLAS method proposed by Arias-Garcia et al. (2021). The KNN-CLAS leverages the voting mechanism of the K-nearest neighbors (KNN) classifier to eliminate the computationally expensive filtering step required during the training phase of NN-CLAS. The original NN-CLAS relies on Gabriel graph computations and a vertex-degree-based noise filtering process to identify structural support edges (SEs) for classification. In contrast, KNN-CLAS bypasses this filtering by aggregating decisions from multiple neighbors, thereby reducing training complexity while maintaining competitive accuracy. Experimental results on benchmark datasets demonstrate that the proposed method achieves comparable performance to NN-CLAS, with significant efficiency improvements. The methodology is particularly suited for embedded systems due to its reduced computational overhead and parameter-free design.**

Index Terms—**pattern recognition, large margin classifiers, Gabriel graph, KNN classifier, embedded systems**

I. INTRODUCTION

Large margin classifiers, such as support vector machines (SVMs), rely on optimization techniques to maximize separation between classes. However, their computational complexity and dependency on user-defined parameters limit their applicability in embedded systems. The NN-CLAS framework, introduced by Torres et al. (2016), addresses these limitations by constructing classifiers directly from the geometric structure of the training data using Gabriel graphs (GGs). The GG encodes pairwise relationships between data points, and support edges (SEs) connecting vertices of opposing classes define local hyperplanes that collectively form a large-margin decision boundary [1]. While effective, NN-CLAS requires a filtering step to remove noisy vertices, which involves evaluating the quality of each vertex based on its neighborhood structure. This process, though critical for robustness, incurs significant computational costs, especially for large datasets [2], [3].

The proposed KNN-CLAS eliminates the need for explicit filtering by leveraging the inherent noise resilience of the KNN voting mechanism. Instead of pruning the dataset during training, KNN-CLAS directly uses the GG's SEs and assigns class labels through a majority vote among the nearest neighbors. This approach retains the structural benefits of GG-based classification while simplifying the training pipeline. The remainder of this section details the original NN-CLAS filtering methodology and its computational challenges.

A. Filtering

The NN-CLAS framework constructs a Gabriel graph G_G from the training set $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}$, where edges connect vertices $\mathbf{x}_i$ and $\mathbf{x}_j$ if no other point lies within the hypersphere defined by their diameter [1]. To handle overlapping classes, a quality measure $q(\mathbf{x}_i)$ evaluates the ratio of same-class neighbors to total neighbors for each vertex:

$$q(\mathbf{x}_i) = \frac{\hat{\mathcal{A}}(\mathbf{x}_i)}{\mathcal{A}(\mathbf{x}_i)},$$

where $\mathcal{A}(\mathbf{x}_i)$ is the vertex degree and $\hat{\mathcal{A}}(\mathbf{x}_i)$ counts neighbors sharing $\mathbf{x}_i$'s class label [2]. Vertices with $q(\mathbf{x}_i)$ below class-specific thresholds t^+ and t^-—calculated as the mean quality per class—are discarded as noise. This filtering ensures SEs lie near the true class boundaries but requires $O(n^2)$ distance computations and iterative quality evaluations, making it impractical for resource-constrained systems [3]. The KNN-CLAS circumvents this bottleneck by integrating neighbor voting, thereby avoiding explicit structural filtering while preserving classification accuracy.

II. METHODOLOGY

III. RESULTS

TABLE I
DATASET METADATA

Dataset	Samples	Features
Ionosphere	351	34
Binary Digits	360	64
Haberman	306	3
Pima Diabetes	768	8
Banknote	1372	4
Sonar	208	60
Breast Cancer	569	30
SPECT Heart	349	44

IV. DISCUSSION

ACKNOWLEDGMENT

REFERENCES

[1] L. C. B. Torres, "Classificador por arestas de suporte (CLAS): métodos de aprendizado baseados em Grafos de Gabriel," Manuscript, 2016.

[2] A. C. Souza, C. Leite Castro, J. A. Garcia, L. C. B. Torres, L. J. Acevedo Jaimes and B. R. A. Jaimes, "Improving the Efficiency of Gabriel Graph-based Classifiers for Hardware-optimized Implementations," 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), Bucaramanga, Colombia, 2019.

TABLE II
DATASET STATISTICS

Dataset	C0/C1	MI	Fisher	Overlap	Imb.Ratio
Ionosphere	0.56	0.21	0.11	0.85	1.79
Binary Digits	0.98	0.18	1.39	0.44	1.02
Haberman	0.36	0.01	0.07	1.0	2.78
Pima Diabetes	1.87	0.05	0.18	0.88	1.87
Banknote	1.25	0.19	0.7	0.75	1.25
Sonar	0.87	0.03	0.09	1.0	1.14
Breast Cancer	0.59	0.21	1.03	0.47	1.68
SPECT Heart	0.37	0.07	0.18	0.91	2.67

TABLE III
MODEL ACCURACY COMPARISON

Dataset	Accuracy			
	nn	1nn	3nn	5nn
Ionosphere	0.88	0.86	0.89	0.88
Binary Digits	1.00	1.00	1.00	1.00
Haberman	0.69	0.69	0.69	0.67
Pima Diabetes	0.75	0.67	0.73	0.75
Banknote	1.00	1.00	1.00	1.00
Sonar	0.76	0.89	0.88	0.88
Breast Cancer	0.95	0.94	0.96	0.96
SPECT Heart	0.68	0.86	0.86	0.82

TABLE IV
TRAINING AND PREDICTION TIMES

Dataset	Training (ms)		Prediction (ms)			
	nn	knn	nn	1nn	3nn	5nn
Ionosphere	115.10	39.20	7.70	4.60	4.90	4.80
Binary Digits	184.60	70.90	2.90	3.30	3.40	3.30
Haberman	14.70	6.70	2.30	2.80	2.50	2.80
Pima Diabetes	98.50	36.70	2.30	4.80	4.50	5.00
Banknote	513.00	91.80	4.90	6.60	6.30	5.90
Sonar	51.50	22.40	2.70	3.00	3.00	2.90
Breast Cancer	160.90	32.40	3.00	3.70	3.70	3.50
SPECT Heart	347.30	150.20	5.90	6.70	6.20	6.30

TABLE V
SUPPORT SAMPLES COUNT

Dataset	Support Samples	
	nn	knn
Ionosphere	111	252
Binary Digits	125	270
Haberman	33	274
Pima Diabetes	141	612
Banknote	170	206
Sonar	89	186
Breast Cancer	101	344
SPECT Heart	98	279

[3] J. Arias-Garcia et al., "Enhancing Performance of Gabriel Graph-Based Classifiers by a Hardware Co-Processor for Embedded System Applications," in IEEE Transactions on Industrial Informatics, vol. 17, no. 2, Feb. 2021.

[4] J. Arias-Garcia et al., "Improved Design for Hardware Implementation of Graph-Based Large Margin Classifiers for Embedded Edge Computing," in IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 1, Jan. 2024.

[5] L. C. B. Torres, C. L. Castro and A. P. Braga, "A parameterless mixture model for large margin classification," 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 2015.

[6] L. C. B. Torres, C. L. Castro, F. Coelho and A. P. Braga, "Large Margin Gaussian Mixture Classifier With a Gabriel Graph Geometric Representation of Data Set Structure," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 3, March 2021.

[7] L. C. B. Torres, C. L. Castro, F. Coelho, F. Sill Torres and A. P. Braga, "Distance-based large margin classifier suitable for integrated circuit implementation," Manuscript, 2015.

[8] D. Dua and C. Graff, "Breast Cancer Wisconsin (Diagnostic) Data Set," UCI Machine Learning Repository, 1995. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

[9] J. Brownlee, "Pima Indians Diabetes Dataset," GitHub Repository, 2020. [Online]. Available: https://github.com/jbrownlee/Datasets

[10] D. Dua and C. Graff, "Haberman's Survival Data Set," UCI Machine Learning Repository, 1995. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Haberman's+Survival

[11] D. Dua and C. Graff, "Data Banknote Authentication Data Set," UCI Machine Learning Repository, 1995. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/banknote+authentication

[12] D. Dua and C. Graff, "Connectionist Bench (Sonar, Mines vs. Rocks) Data Set," UCI Machine Learning Repository, 1995. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/connectionist+bench+(sonar,+mines+vs.+rocks)

[13] D. Dua and C. Graff, "Adult Data Set," UCI Machine Learning Repository, 1995. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/adult

[14] D. Dua and C. Graff, "Ionosphere Data Set," UCI Machine Learning Repository, 1995. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/ionosphere

[15] D. Dua and C. Graff, "SPECT Heart Data Set," UCI Machine Learning Repository, 1995. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/SPECT+Heart

[16] L. Breiman et al., "Optical Recognition of Handwritten Digits Data Set," Scikit-learn Documentation, 1998. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_digits.html