International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Classification of Imbalanced Data Using a Modified Fuzzy-Neighbor Weighted Approach

Patel Harshita[1]* **Thakur Ghanshyam Singh [2]**

[1, 2] *Department of Mathematics and Computer Applications, Maulana Azad National Institute of Technology, Bhopal-462003, India*
* Corresponding author's Email: hpatel.sati@gmail.com

Abstract: Classification of imbalanced datasets is one of the widely explored challenges of the decade. The imbalance occurs in many real world datasets due to uneven distribution of data into classes, i.e. one class has more instances while others have a few that results in the biased performances of traditional classifiers towards the majority class with large number of instances and ignorance of other classes with less data. Many solutions have been proposed to deal with this issue in various crisp and fuzzy methods. This paper proposes a new hybrid fuzzy weighted nearest neighbor approach to find better overall classification performance for both minority and majority classes of imbalanced data. Benefits of neighbor weighted K nearest neighbor approach i.e. assignment of large weights to small classes and small weights to large classes are merged with fuzzy logic. Fuzzy classification helps in classifying objects more adequately as it determines that how much an object belongs to a class. Experimental results exhibit the improvements in classification of imbalanced data of different imbalance ratios in comparison with other methods.

Keywords: Imbalanced data, K nearest neighbor, Fuzzy K nearest neighbor, Classification.

1. Introduction

In today's fast pace scientific world, data generation is accelerated with exponential speed with availability of highly techno-friendly equipment from almost all areas of society. This brings various data mining opportunities with many challenges [1, 2]. Classification of imbalanced data has drawn significant attention of researchers from data mining and machine learning community. The issue got the interest due to existence of imbalance in real world datasets and inability of traditional classifiers to classify it correctly. We can understand the term with unequal or uneven distribution of data into classes. One class with more instances is known as majority class and other with fewer instances as minority class. Majority class with large number of instances dominate the classification results and minority classes are misclassified, even most of the time they are of most interest. The 10 challenging problems in data mining research includes the learning from imbalanced data [3] and this issue is coming under new trends of data mining [4]. Raeder mentioned it mentioned it specifically again in [5]. Examples of real world imbalanced datasets are Medical Diagnosis [6, 7], Oil-spill Detection [8], Credit Card Fraud Detection, Network Intrusion, Text Categorization, Helicopter Gearbox Fault Monitoring [9] etc.

Four popular solutions are available to deal with this issue are (a) balance the data by re-sampling, (b) modification in traditional classification algorithms, (c) cost sensitive methods and (d) ensemble approaches [9]. In this paper we adopt second solution of modification of traditional classification algorithm and we study the nearest neighbor approach with its weighted and fuzzy variant to find proper classification results of imbalanced data. K nearest neighbor is a one of the top data mining algorithms [10] that is easily understandable, robust, and easily programmable. Error rate of K nearest neighbor is bounded above by twice the Bayes error

rate [11, 12]. The proposed method gives improved version of neighbor weighted K nearest neighbor with combination of fuzzy logic. Weights from neighbor weighted K nearest neighbor algorithm become stronger with the use of fuzzy concepts by getting opportunity of classification on soft boundary values instead of crisp's hard calculations. The membership function we used in this paper is a basic membership function given by Keller et. al. [26] designed for binary fuzzy classification. It is a benchmark membership function used and applicable in many nearest neighbor classification tasks. In results we can observe the better performance of proposed approach on different evaluation measures over the neighbor weighted and other algorithm.

The organization of the paper is as follows: Related work is shown in second section of paper which is followed by preliminaries in section three. Fourth section is dedicated for our proposed method. Section five contains experimental and results for the proposed approach and the last sixth section is providing concluding remarks with future scope.

2. Related Work

Conventional classification strategies assume the quantity of data into all classes is equal by default, so they are insensitive to the imbalanced distribution of the data. In results they lead to misclassification for the small classes having few elements in comparison with the large classes. In other proposed solutions re-sampling is very popular. Though it presents good classification of imbalanced data but in some manner it affects original distribution of data such as duplication of information in oversampling and data loss (information loss) in under-sampling techniques [9]. Cost sensitive approaches are also not applicable where cost is not given. Algorithm modification with almost all available classification algorithms is done by many researchers for imbalanced data. Modified K nearest neighbor approaches provide different solutions to deal with unequally distributed data.

This section includes related literature with the proposed concept and providing knowledge of the field. A single CCNND algorithm is proposed by Kriminger et. al. [13], they applied local geometric structure in data. This approach is applicable on multi class imbalance as well as allows classification for different degrees of imbalance. Tomasev et. al.[14] have worked on K nearest neighbor's hubness effect on high dimensional data which is responsible for high misclassification. This misclassification occurs due to minority classes while in low and medium dimensional datasets misclassification occur due to majority classes. For cross project defect prediction (CPDP) Ryu et. al. [15] proposed an instance hybrid selection using nearest neighbor (HISNN). In such cases class imbalance exists in source and target project distribution.

Many weighted KNN are performing well on imbalanced datasets. Some are being discussed here. Liu et. al. [16] proposed the concept of class confident weights to sort out the imbalance issue of data. They found posterior probabilities by converting attribute prior probabilities to weight prototypes. Class based weighted nearest neighbor algorithm is proposed by Dubey et. al. [17] for imbalanced data classification. The calculation of weights was based on the distribution of nearest neighbor of unknown examples. Patel et. al. [18] proposed a hybrid neighbor weighted approach to deal with large and small weight assignment to minority and majority classes with large and small K respectively. A mathematical model was proposed by Ando [19] with convex optimization technique by given to learn class wise weighted nearest neighbor approach that maximize nonlinear performance measure for training data.

Beside all these practices, fuzzy solutions too have a good scope to cope up with imbalance problem. A few contributions are being discussed here. Fernandez et. al. [20] analyzed fuzzy rule based system for imbalanced datasets. For improved classification of different imbalance ratios they applied adaptive parametric conjunction operators. Han et. al. [21] proposed a nearest neighbor approach with fuzzy and rough properties to minimize a biasness generated due to majority class. To solve imbalance problem in categorical data Liu et. al. [22] proposed a coupled fuzzy K nearest neighbor approach for unequally distributed data with strong bonds among attribute, classes and instances. Important functions of this approach are assignment of sized membership, similarity calculation and integration. Ramentol et. al. [23] proposed a fuzzy rough weighted nearest neighbor algorithm for imbalanced data, they proposed six weight vectors and some indiscernibility relations with these weight vectors. Our proposed method is inspired from these works and giving an improved easily computable neighbor weighted solution with fuzzy K nearest neighbor.

3. Preliminaries

Basic knowledge of K-nearest neighbor algorithm [12][24][25], fuzzy K-nearest neighbor

algorithm [26] and neighbor weighted approach [27] help to understand better the proposed method. Euclidian distance is considered as distance measure for all nearest neighbors discussed here in this paper by default. Mathematical models for these preliminaries are explained in following subsections.

3.1 K Nearest Neighbor

The K-nearest neighbor algorithm finds nearest neighbors of an unknown instance q_u from training dataset on the basis of distance between q_u and all training instances then class label is assigned to q_u for the class having maximum neighbors.

Mathematical formulation could be understood in following equation:

$$C(q_u) = \arg\max_{C \in \{Cj | j=1,2\}} \sum_{xi \in S(q,K)} T(x_i, C) \tag{1}$$

Here $C(q_u)$ = *class label of* q_u, *to be predicted*,
m = *Number of classes*,
$S(q_u, K)$ = *Set of* $K-$*nearest neighbors of* q_u *and*

$$T(x_i, C) = \begin{cases} 1 & \text{if } x_i \in C \\ 0 & \text{otherwise} \end{cases}$$

3.2 Neighbor Weighted K nearest Neighbor

Neighbor Weighted K nearest Neighbor approach for imbalanced text datasets is proposed by Tan [27] for. Idea behind this method is to assign large weights to small classes and small weights to large classes to minimize the biasness of the classifier headed for majority class and avoidance of minority class. These weights help in more accurate classification of imbalance data.

After finding K nearest neighbors for query instance q_u from traditional K nearest neighbor method we find weights from following equation:

$$W_j = \frac{1}{(N(C_j) / Min\{N(C_j) \mid j = 1,2\})^{1/p}} \tag{2}$$

p is an exponent and p>1,

Here W_j = *Weight for* j^{th}*class, to be predicted*,
$N(C_j)$ = *Number of nearest neighbors of* q_u
belongs to class j

These weights are combined with traditional classification algorithm to find class label of a query instance q_u.

$$C(q_u) = \arg\max_{C \in \{Cj | j=1,2\}} W_i \left(\sum_{xi \in S(q_u, K)} T(x_i, C) \right) \tag{3}$$

3.3 Fuzzy K-Nearest Neighbor Algorithm

The K-nearest neighbor algorithm in fuzzy form finds memberships of data examples into classes instead of finding their complete belongingness. This improves the accuracy of classification.

Keller et. al. (1985) [26] proposed the following model for fuzzy memberships of training data instances into classes.

if $x \in C$ *and* $C = m$, *Then*

$$\mu_c(x) = \begin{cases} 0.51 + (n_c / K) * 0.49 & \text{If } C = m \\ (n_c / K) * 0.49 & \text{othrerwise} \end{cases} \tag{4}$$

Here n_C = *nearest neighbors of* x *from class* C
$\mu_C(x)$ = *Membership of* x *into class*C

And for memberships of test instance q

$$\mu_c(q_u) = \frac{\sum_{i=1}^{K} \mu_{ci} (1 / \|q_u - q_i\|^{2/(p-1)})}{\sum_{i=1}^{K} (1 / \|q_u - q_i\|^{2/(p-1)})} \tag{5}$$

Where p *is an integer and* $p > 1$
And q_i *is nearest neighbor of* q_u, $(i = 1....K)$

4. Proposed Method

The proposed approach is a fuzzy extension of neighbor weighted nearest neighbor method [27]. Neighbor weighted approach (further called as NWKNN in this paper) is a good weighting strategy for imbalanced data classification that was proposed for text data and more accurately classify the imbalanced data than traditional nearest neighbor approach by assigning large weights to small classes and vice versa. With fuzzy logic, performance of neighbor weighted approach improved as membership says that how much an instance belongs to a class instead of assigning whole class label in hard manner. We first find memberships of

all training instances in step one. Then in step two find out the nearest neighbors for query instance using traditional K nearest neighbor approach. Weights will be calculated for all classes in step three. In step four we find the membership of query instance and class label is assigned in step five on the basis of membership of query instance in all classes.

Algorithm (Fuzzy-NWKNN)

Input: Training Dataset *D*, Query instance x_t, Set of class labels *C* and Parameter *K*.

Output: Class label of x_t.

Step 1. Find memberships of training data into each class using

$$\text{Let } x \in C_m \text{ ,Then}$$

$$\mu_{Ci}(x) = \begin{cases} 0.51 + (n_{Ci}/K) * 0.49 & If \ \ i = m \\ (n_{Ci}/K) * 0.49 & otherwise \end{cases}$$

$$While\ taking \ \sum \mu_{Ci}(x) = 1$$

Step 2. Find K nearest Neighbor for x_t

Step 3. Obtain weights with

$$W_j = \frac{1}{(N(C_j)/Min\{N(C_j) \mid j = 1,2\})^{1/p}}$$

Step 4. Find class memberships of test instance x_t using following formula

$$\mu_{ci}(x_t) = \frac{\sum_{j=1}^{K} w_j * \mu_{cij}}{\sum_{j=1}^{K} w_j}$$

$$While\ taking \ \sum \mu_C(x_t) = 1$$

Step 5. Assign class label to test instance x_t by

$$C_m(x_t) = \begin{cases} C_m & if\ \mu_c(x_t) \geq 0.51 \\ Random\ Assignment & Otherwise \end{cases}$$

5. Experiments and Results

Experiments for proposed method are done for binary imbalanced datasets. All features are taken for classification, keeping into the mind that all features of datasets playing important role in their classification no specific feature selection method is applied. Also it becomes very ambiguous for imbalanced data to decide which attribute is of more interest. Following subsections are providing description of datasets, about evaluation measure and comparative outcomes of proposed method with its predecessor.

5.1 Datasets

Six datasets are taken from UCI [28] and KEEL [29] repositories. These datasets are of different natures like disease data, glass quality or radar signal's data, all are imbalanced with uneven distribution of instances into classes. Brief description of data is given in table 1.

Table 1. Short Description of Datasets

Datasets	Source	# Instances	Class (1/0)	# Attributes	Imbalance Ratio
Ionosphere	UCI	351	Bad/Good Radar Returns	34	1.79
Wisconsin	KEEL	683	Positive/Negative	9	1.86
Phoneme	KEEL	5404	Oral/Nasal Sound	5	2.4
Vehicle0	KEEL	846	Positive/Negative	18	3.3
New-Thyroid	UCI	215	Positive/Negative	5	5.14
Glass2	KEEL	214	Positive/Negative	9	11.6

5.2 Evaluation Criteria

Accuracy is a common evaluation measure used to judge the performance of a classifier. But this seems inadequate while dealing with imbalanced datasets. F-measure, AUC and G-mean are some of the popular measures for evaluation of classifiers for such datasets. In this paper we are evaluating and comparing the performance of our proposed method

with neighbor weighted approach on these evaluation criteria. To calculate the results of these measures confusion metric is needed, given as follows:

Table 2. Confusion Metric for Binary Classification

Predictive/ Actual Outcomes	Predicted Positive	Predicted Negative
Actual Positive	True Positive	False Positive
Actual Negative	False Negative	True Negative

True positive (TP): Actual positive instances correctly classified as positive,

False positive (FP): Actual positive instances incorrectly classified as negative,

True negatives (TN): Actual negative instances correctly classified as negative,

False negatives (FN): Actual negative instances incorrectly classified as positive.

Based on confusion metric evaluation measures could be understood in terms of following equations:

$$\Pr esicion = \frac{TP}{TP + FP} \tag{6}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{7}$$

$$TP_{Rate} = \frac{TP}{Total\_P} \tag{8}$$

$$FP_{Rate} = \frac{FP}{Total\_N} \tag{9}$$

$$F - Measure = \frac{2 * \Pr ecision * \mathrm{Re}\,call}{\Pr ecision + \mathrm{Re}\,call} \tag{10}$$

$$AUC = \frac{1 + TP_{Rate} - FP_{Rate}}{2} \tag{11}$$

$$G - Mean = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}} \tag{12}$$

5.3 Empirical Results

Evaluation of our proposed method (Fuzzy-NWKNN) on all six datasets described in Table 1 is performed with F-measure, AUC and G-mean and its performance is compared with neighbor weighted algorithm (NWKNN) and Hybrid weighted nearest neighbor approach (Adpt-NWKNN) [18] for imbalanced data. Results for all mentioned evaluation measures are shown from Table 3 to Table 8 for individual datasets. We performed experiment for five values of K to find more generalized results.

Table 3. Results for F-Measure, AUC and G-Mean for Ionosphere dataset

K	F-Measure			AUC			G-Mean		
	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN
5	0.4252	0.4252	0.4733	0.3926	0.3926	0.4452	0.2303	0.2303	0.2467
10	0.4375	0.4733	0.4615	0.4057	0.4452	0.4321	0.2345	0.2467	0.2427
15	0.4561	0.4957	0.5085	0.469	0.5084	0.5216	0.4167	0.44	0.4476
20	0.52	0.549	0.5631	0.5734	0.5998	0.6129	0.5626	0.5839	0.5942
25	0.5	0.5769	0.5743	0.5621	0.6261	0.6278	0.5577	0.6044	0.6131

Table 4. Results for F-Measure, AUC and G-Mean for Wisconsin dataset

K	F-Measure			AUC			G-Mean		
	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN
5	0.5147	0.5147	0.7609	0.4974	0.4974	0.8282	0.1481	0.1481	0.8156
10	0.5259	0.5221	0.9272	0.5156	0.5081	0.9523	0.2109	0.1722	0.9521
15	0.5299	0.5279	0.9655	0.5231	0.5194	0.9748	0.2435	0.2278	0.9748
20	0.5243	0.5299	0.9655	0.5162	0.5231	0.9748	0.2418	0.2435	0.9748
25	0.5188	0.5259	0.9655	0.5092	0.5156	0.9748	0.2401	0.2109	0.9748

Table 5. Results for F-Measure, AUC and G-Mean for Phoneme dataset

K	F-Measure			AUC			G-Mean		
	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN
5	0.4236	0.4236	0.4511	0.4936	0.4936	0.5278	0.3993	0.3993	0.4237
10	0.4476	0.4593	0.476	0.5334	0.5417	0.5716	0.4676	0.4506	0.5131
15	0.443	0.491	0.5016	0.5432	0.5842	0.6083	0.5155	0.4974	0.5689
20	0.4574	0.5038	0.5169	0.5627	0.6024	0.6286	0.5394	0.5243	0.5983
25	0.4587	0.5216	0.5313	0.5712	0.6255	0.6465	0.5589	0.5495	0.6224

Table 6. Results for F-Measure, AUC and G-Mean for Vehicle0 dataset

K	F-Measure			AUC			G-Mean		
	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN
5	0.3491	0.381	0.3732	0.4696	0.5079	0.5009	0.3337	0.2774	0.3236
10	0.3698	0.4096	0.3955	0.5063	0.5541	0.5422	0.4	0.329	0.4214
15	0.3465	0.4	0.4015	0.4801	0.5381	0.5525	0.4078	0.3021	0.4425
20	0.3629	0.4152	0.4157	0.5064	0.5644	0.5757	0.444	0.359	0.4866
25	0.339	0.4138	0.424	0.4828	0.5619	0.5886	0.4464	0.3517	0.5094

Table 7. Results for F-Measure, AUC and G-Mean for New-Thyroid dataset

K	F-Measure			AUC			G-Mean		
	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN
5	0.2388	0.2609	0.2647	0.4463	0.487	0.4963	0.2722	0.2582	0.2887
10	0.2373	0.2727	0.2951	0.4611	0.5148	0.5611	0.3944	0.3416	0.4472
15	0.2456	0.2687	0.3051	0.4796	0.5056	0.5796	0.426	0.3162	0.483
20	0.25	0.2857	0.3103	0.4889	0.5426	0.5889	0.441	0.4082	0.5
25	0.2545	0.2857	0.3158	0.4981	0.5426	0.5981	0.4554	0.4082	0.5164

Table 8. Results for F-Measure, AUC and G-Mean for Glass2 dataset

K	F-Measure			AUC			G-Mean		
	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN	NWKNN	Adpt-NWKNN	Fuzzy-NWKNN
5	0.1	0.0938	0.1311	0.3593	0.3254	0.4593	0.2668	0.1747	0.3081
10	0.0408	0.1449	0.1154	0.2356	0.5	0.4271	0.2329	0	0.3906
15	0.0465	0.127	0.1277	0.2864	0.4424	0.4695	0.2731	0.2604	0.451
20	0.0541	0.1667	0.1463	0.3373	0.5763	0.5203	0.3081	0.3906	0.5142
25	0.0606	0.1587	0.1622	0.3712	0.5508	0.5542	0.3294	0.3189	0.5523

Comparison of performances could also be represented in form of bar graphs. Figure 1 to figure 3 are showing the combined result for all datasets for F-measure, AUC and G-mean for average values of K.

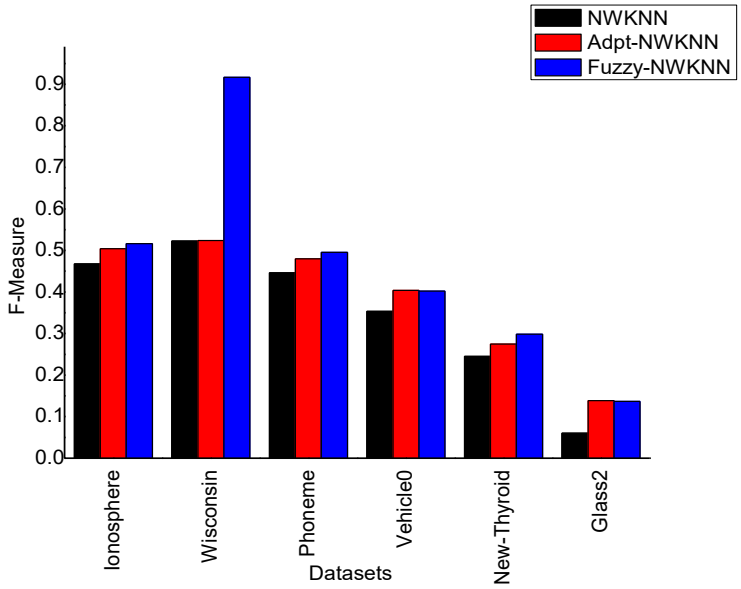

Figure 1. F-measure Performances for NWKNN, Adpt-NWKNN and Fuzzy-NWKNN

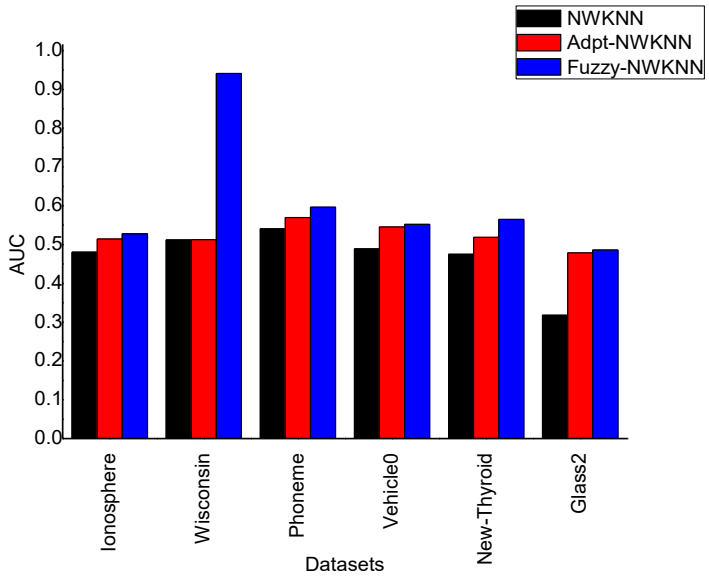

Figure 2. AUC Performances for NWKNN, Adpt-NWKNN and Fuzzy-NWKNN

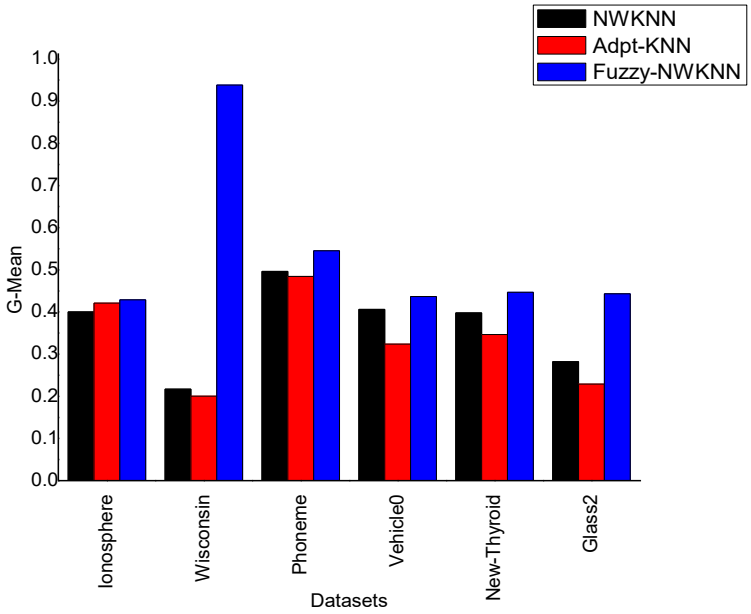

Figure 3. G-Mean Performances for NWKNN, Adpt-NWKNN and Fuzzy-NWKNN

6. Conclusion and Future Scope

Proposed fuzzy neighbor weighted algorithm is a fruitful extension of neighbor weighted approach. Neighbor weighted approach was performing well with imbalanced text data. In this paper we extract results of neighbor weighted method for numerical imbalanced data and proposed its fuzzy variant. In comparative study our proposed method (Fuzzy-KNN) is extracting better results than NWKNN and Adpt-NWKNN for all three evaluation measures on datasets from table 1. We consider all features necessary for classification. In further studies feature selection could also be considered. The experiments were performed for binary class datasets, this study could be extended for multiclass datasets in future.

References

[1] J. Han and M. Kamber, *Data Mining, Concepts and Techniques*, Morgan, Kaufmann, 2000.

[2] H. Patel and D.S. Rajput, "Data Mining Applications in Present Scenario: A review", *International Journal of Soft Computing*, Vol. 6, pp. 136-142, 2011.

[3] Q. Yang and X. Wu, "10 challenging problems in data mining research", *International Journal of Information Technology and Decision Making*, Vol. 5, pp. 597-604, 2006.

[4] Editorial, Special issue on "New trends in data mining", NTDM. *Knowledge Based Systems*, Elsevier, pp. 1-2, 2012.

[5] T. Raeder, G. Forman and N. V. Chawla, "Learning from Imbalanced Data: Evaluation Matters", *in D.E. Holmes, L.C. Jain (Eds). Data Mining: Foundations and Intelligent Paradigms*, ISRL 23, pp. 315-331, 2012.

[6] R. Pavón, R. Laza, M. Reboiro-Jato and F. Fdez-Riverola, "Assessing the impact of class-imbalanced data for classifying relevant/irrelevant medline documents", *Advances in Intelligent and Soft Computing*, Vol. 93, pp. 345–353, 2011.

[7] R. B. Rao, S. Krishanan and R.S. Niculscu, "Data Mining for Improved Cardiac Care", *ACM SIGKDD Exploration Newsletter*, Vol. 8, pp. 3–10, 2006.

[8] M. Kubat, R. C. Holte and S. Matwin, "Machine Learning for the Detection of Oil Spills in Satellite Images", *Machine Learning*, pp. 195–215, 1998.

[9] H. He and E. A. Garcia, "Learning from Imbalanced Data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21, pp. 1263-1284, 2009.

[10] X. Wu et al., "Top 10 Algorithms in Data Mining", *Knowledge Information Systems*, Vol. 14, pp. 1-37, 2008.

[11] G. Loizou and S. J. Maybank, "The Nearest Neighbor and the Bayes Error Rates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, pp. 254-262, 1987.

[12] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification", *IEEE Transactions on Information Theory*, Vol. 13, pp. 21-27, 1967.

[13] E. Kriminger and C. Lakshminarayan. "Nearest Neighbor Distributions for Imbalanced

Classification", In: *Proc. of WCCI 2012 IEEE World Congress on Computational Intelligence*, Brisbane, 2012, pp. 10-15.

[14] N. Tomašev and D. Mladenic. "Class Imbalance and the Curse of Minority Hubs", *Knowledge-Based Systems*, Vol. 53, pp. 157–172, 2013.

[15] D. Ryu, J. Jang and J. Baik, "A hybrid instance selection using nearest-neighbor for cross-project defect prediction", *Journal of Computer Science and Technology*, Vol. 30, pp. 969-980, 2015.

[16] W. Liu and S. Chawla. "Class Confidence Weighted kNN Algorithms for Imbalanced Data Sets", *PAKDD 2011, Part II, LANI 6635*, pp. 345-356, 2011.

[17] H. Dubey and V. Pudi. "Class based weighted k nearest neighbor over imbalanced dataset", *PAKDD 2013, Part II, LANI, 7819*, pp. 305-316, 2013.

[18] H. Patel and G.S. Thakur, "A Hybrid Weighted Nearest Neighbor Approach to Mine Imbalanced Data", *In: Proc. of the 12th International Conference on Data Mining (DMIN'16),* pp 106-110, 2016.

[19] S. Ando, "Classifying imbalanced data in distance-based feature space", *Knowledge and Information Systems*, vol. 46, pp. 707–730, 2016.

[20] A. Fernandez, M. J. Jesus and F. Herrera, "On the Influence of an Adaptive Inference System in Fuzzy Rule Based Classification Systems for Imbalanced Data-Sets", *Expert Systems with Applications*, Vol. 36, pp. 9805-9812, 2009.

[21] H. Han and B. Mao, "Fuzzy-Rough k-Nearest Neighbor Algorithm for Imbalanced Data Sets Learning", In: *Proc. of FSKD 2010-Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, IEEE circuits and systems society, China, pp. 1286-1290, 2010.

[22] C. Liu, L. Cao and P.S. Yu, "Coupled Fuzzy *K*-Nearest Neighbors Classification of Imbalanced Non-IID Categorical Data", In: *Proc. of IJCNN - International Joint Conference on Neural Networks*, IEEE, Beijing, 2014, pp. 1122-1129.

[23] E. Ramentol, S. Vluymans, N. Verbiest, Y. Caballero, R. Bello, C. Cornelis, and F. Herrera, "IFROWANN: Imbalanced Fuzzy-Rough Ordered Weighted Average Nearest Neighbor Classification", *IEEE Transactions on Fuzzy Systems*. 2014.

[24] E. Fix and J. L. Hodges, "Discriminatory Analysis-Nonparametric Discrimination: Consistency Properties", *Technical Report 4, Project no. 21-29-004, USAF School of Aviation Medicine, Randolph Field*, Texas, 1951.

[25] E. Fix and J. L. Hodges, "Discriminatory Analysis-Nonparametric Discrimination: Consistency Properties", *International Statistical Review*, Vol. 57, pp. 238–247, 1989.

[26] J. M. Keller, M. R. Grey and J. A. Givens Jr., "A Fuzzy k- Nearest Neighbor Algorithm", *IEEE Transactions on System, Man and Cybernetics*, Vol. 15, pp. 580-585, 1985.

[27] S. Tan, "Neighbor-weighted K-Nearest Neighbor for unbalanced text corpus", *Expert Systems with Applications*, vol. 28, no. 4, pp. 667–671, 2005.

[28] A. Asuncion and D. J. Newman, UCI machine learning repository.
University of California, School of Information and Computer Science, Irvine, 2007.
http://www.ics.uci.edu/~mlearn/MLRepository.html

[29] KEEL: *Knowledge Extraction based on Evolutionary Learning*.
http://sci2s.ugr.es/keel/imbalanced.php