JCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-10, Oct 2018 E-ISSN: 2347-2693

FUZZY GRAVITATIONAL CLASSIFIER FOR CLASSIFYING IMBALANCED DATASETS

Sinciya P.O.[1*], J. Jeya A Celin[2]

[1]Dept. of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India
[2]Dept. of Information technology, Noorul Islam University, Kumaracoil, India

[*]*Corresponding Author: sinciyapo@gmail.com, Tel.: 7598221021*

Available online at: www.ijcseonline.org

Accepted: 20/Oct/2018, Published: 31/Oct/2018

Abstract— Developing a precise and consistent model for classifying imbalanced medical data is one of the major challenges in machine learning and data mining. As the advanced growth in medical technology, a classy medical classification system is essential that make use of data mining algorithms to support medical diagnosis practice. Though the standard medical data seldom obeys the requirements of different knowledge engineering tools, most of the medical datasets are considered to be highly imbalanced with respect to their class label. So the imbalancing problem has been found to thwart the efficiency of the learning model. The only way to avoid this problem is to reduce the gap between both majority and minority class instances. In our approach a fuzzy gravitational classifier with weighting scheme is employed, in which weight is optimized using Particle swarm optimization algorithm. The technique is implemented and tested with three well known bench mark imbalanced dataset from UCI and KEEL repository. A comparative study is made with two existing classification methods viz. Weighted nearest neighbour and class based weighted nearest neighbour. Evaluation results shows our hybrid approach gives better performance on imbalanced data in terms of AUC, F-measure and G-mean.

Keywords— Imbalanced data, PSO optimization, Data gravitation classifier, Fuzzy soft set, JFIM

I. INTRODUCTION

In recent year's imbalanced classification have attracted an important research problem in machine learning and data mining. In standard data classification, the dataset contains equal distribution of instances in each class. Unlike this, imbalanced classification has an imbalanced class distribution. It may be binary class imbalancing or multi class imbalancing. Binary class imbalancing means the number of samples in one class called majority class have more number of samples than in another called minority class. On the other hand, multi class imbalancing means one or more classes might be outnumbered by the number of samples while the other classes contain only a few. Standard classification algorithms are not appropriate to handle imbalancing problems because these models are over trained by majority class instances. Although standard classifier gives high classification performance across the entire dataset, the actual performance is very low because its classifying rate of minority class is low. The features which affect the classification performance of imbalanced data are small disjuncts [1] [2], class overlaps [3], and noise [4], and borderline samples [5].

Most of the real world applications such as Internet traffic problems [6], fraud detection [7] [8], text processing [9], remote sensing [10], intrusion detection [11] [12], fault identification [13], image processing [14] and medical diagnosis [15] have related identification features. In all these cases, the minority class has a major role than the majority class [16, 17]. Thus giving maximum importance to the minority class is the main goal of an imbalanced classification task.

In imbalanced binary classification tasks, the dominant class is called positive class and subordinate class is named as negative class. The basic measure of evaluating imbalancing is the Imbalance ratio (IMR). It is defined as the ratio of number of samples in positive and negative class.

$$IMR = \frac{P}{N}$$

Where P is the number of samples in positive class and N is the number of samples in negative class.
The content of the paper is organized as follows. Section 2 describes the related literature review. Section 3 & 4 gives the idea of fuzzy soft set and gravitational classifier. Section 5 presents the proposed Fuzzy gravitational classifier for imbalanced datasets. Section 6 outlines the evaluation criteria and conducted experimental results.

II. RELATED WORK

Dudani proposed a distance weighted NN (DWKNN) [18] rule to assign a high weight to the instance which are closer to the unknown sample and less weight to the instance which

© 2018, IJCSE All Rights Reserved

are at greater distance from the unknown sample. Consider $i_1, i_2, .., i_k$ are the set of instances and $d_1, d_2, .., d_k$ are the corresponding distances from the unknown sample, form which i_1 is the closest and i_k is the furthest instance. Then the weight for the instance i_j is defined as:

$$\mathrm{Wj} = \begin{Bmatrix} \dfrac{d_k - d_j}{d_k - d_1}, & if\ d_k \neq d_1 \\ 1 \quad , & if\ d_k = d_1 \end{Bmatrix}$$

DW-KNN assigns the unknown sample to the class, which gives the highest weighted sum value.

Gao and Wang introduced a center-based NN classifier (CNN) [19] which finds the distance between centre of their class and training instances and finds how far the training sample from unlabelled sample. This method does not perform well if the cenres of their data classes are overlapped.

Wang *et al.* have designed an adaptive *k* NN algorithm (KNN-A) [20] which used an adaptive distance measure and weighting scheme to correctly classify the test instances.
Paredes and Vidal [21], [22] proposed a class-dependent weighted distance measure in to enhance the efficiency of the NN classifier. This approach classifies the data sample by considering distance between points in the same class is small and in different class is large. They extend their algorithm with prototype reduction algorithm, which starts with small number of prototypes and iteratively adjust the position and weight matrix of the features.

Peng *et al.* [23] introduced a data gravitation classifier which employs weights to the features in distance calculation. The weights are optimized by a random iterative algorithm named tentative random feature selection (TRFS). The gravitation for each particle is calculated by mass and distance from centroid. This method gives better performance in terms of accuracy, F-measure and Recall. But it does not handle imbalanced data. In literature, most of weighted approaches uses nearest neighbour algorithm for classification. In this work, we have developed a hybrid approach which overcomes the problem of imbalancing by combining fuzzy soft set with gravitational classifier.

III. FUZZY SOFT SET

Many classification algorithms proved that the use fuzzy set theory is a good choice for dealing with uncertainties. But there is no suitable mechanism to deal with membership function because it may change on the basis of problem domain. In order to avoid these problems, a new mathematical model is introduced by Molodtsov called soft set concept which have necessary parameterization to deal uncertainty problems.

Definition 1: *Soft set.* Suppose U be an initial universal set and T be a set of features. Let P(U) be the set of all subset or power set of U and $A \subset T$. A pair vector (F, A) is termed as soft set over U where F is a mapping given by $F: A \rightarrow P(U)$.
Definition 2: *Fuzzy Soft set.* The family of all fuzzy sets of U is denoted by Ƒ (U). Let $A_i \subseteq E$. Then a pair (F_i, A_i) is called a fuzzy soft set over U, where F_i is a mapping given by $F_i: A_i \rightarrow$ Ƒ (U).

A. Comparison table

Comparison table is a square table with the number of rows equal to the number of columns, and both the rows and the columns are labelled by class names c*1*, *c2… cn* of the universe. Each value in the table (*gij*) = the number of attributes of class c*i* that is greater than or equal to the attributes of class c*j* in terms of the membership value. Therefore, $0 \le gij \le p$, and $gii = p$, $\forall i, j$, where *p* is the number of parameters in a fuzzy soft set. Thus, *gij* is a numerical measure, which is an integer value, and c*i* dominates c*j* in *gij* number of parameters out of *p* parameters.
Row sum for each class c*i*, sum of weights of c*i* that are greater than other class weights, can be calculated using Equation (3).

$$r_i = \sum_{j=1}^{n} g_{ij}$$ □□□

Thus, *ri* indicates the total number of parameters in which c*i* dominates all the members of *U*. Column sum of an class c*i*, sum of weights of other instances that are greater than c*i* can be calculated as below.

$$c_i = \sum_{i=1}^{n} g_{ij} \qquad (2)$$

The value *ci* indicates the total number of parameters in which c*j* is dominated by all the members of *U*. Score of an instance *i* is calculated by the difference of row sum and column sum is given by Equation (1) & (2).

$$S_i = r_i - c_i, i = 1,2 \ldots n. \quad (3)$$

Based on this score value, the unknown class label of an instance can be determined.

IV. GRAVITATIONAL CLASSIFICATION

A lot of classification and clustering algorithms are suggested on the notion of distance between data instances. The most commonly used distance measure to find similarity of two data sample is the Euclidean distance. When we analyze the association of a test sample and a training data set in a data cluster or in a data class, two important factors are considered. First one is the number of instances and the other is the distance. The test instance is assigned to the class in which the data set contains more number of instances or the instance in the group is at shorter distance with respect to the class. In this paper we extract the similarity of two data samples in terms of gravitation, which is defined by distance and data mass.

Definition 1: (Data Unit). Data Unit is a collection of data particles in the data space. All these data elements possess some kind of relationship. Normally this relationship denotes the distance between data elements. The distance

© 2018, IJCSE All Rights Reserved

between any two data particle in the data unit should be lesser than already available value. The data particles in the data unit have a data mass and centroid.

Definition 2: (Data mass). Data mass is the number of data particles in the data unit.

Lemma 1: Assume the training data set containing two classes C1 and C2.For an unknown data T, the data samples in C1 acts on T is denoted as gravitational force F1 and data samples in C2 acts on T is denoted as F2. Assign T to Class C1 if the gravitation force in C1 is greater than C2. Ie. F1>F2.

Consider a set of training samples S={t_1,t_2,…t_n} in an n-dimensional space, where t_1=(x_1,y_1), t_2=(x_2,y_2) and t_n=(x_n,y_n), C={c_1,c_2,…c_d} represents d number of classes and n is the total number of training instances. For a given test data considered as atomic value, denoted by D and x is its centroid position.

The gravitational force for each class j on data particle D is given by

$$F_j = \sum_{k=1}^{d} \frac{m_{jk}}{\left|x_{jk} - x\right|^2} \qquad (5)$$

Where m_{jk} is the mass of the data sample j in class k and its mass centre is x_{jk}.

The main idea of gravitational classification is to compute gravitational force for all data classes. According to lemma 1, assign test data to the class which possess maximum gravitation.

But the gravitational classifier does not give better solution for the imbalancing problem. To avoid this issue, in our approach we have used a class dependent weight matrix denoted as W_f[N, c] with fuzzy approach, where N is the number of features and C represents the class.

$$W_f = \begin{bmatrix} w_{1,1} & \cdots & w_{1,c} \\ \cdots & \cdots & \cdots \\ w_{N,1} & \cdots & w_{N,c} \end{bmatrix} \qquad (6)$$

Therefore, when weight is added to distance function it becomes weighted distance, and it's given by $d(X, Y, c) = \sqrt{\sum_{i=1}^{N} W_f \cdot (X_i - Y_i)^2}$ (7)

Where W_f is the weight of feature *i to* the corresponding class c. The optimized set of weight is calculated by weight optimized particle swarm algorithm(WOPS)[24].

V. PROPOSED FUZZY GRAVITATIONAL CLASSIFIER

Step1: Calculate the centroid vector *Ci* (for each class *i)*, by calculating the average value of data present in the dataset *Di* (*Di* set of instances of class *i*) using Equation

$$\vec{C_i} = \frac{1}{D_i} \sum_{d_j \in D_i} \vec{d_j} \qquad (4)$$

Step2: Represent the centroid vectors as a table of size *I X N* (*I* classes and N features) which can be considered as a soft set (F, E). An entry in the table is *gin*, *i* = 1, 2, ...,*I* and *n* = 1, 2, ...,*N* .

Step 3: Get a feature vector *Ef* from the unknown dataset.

Step 4: Generate a soft set (F, A) with its entry as *kin*, *i* = 1, 2… *I* and n= 1, 2... *N*, calculated using the formula,

$$kin = 1 - \left\{ \left(\frac{N_c}{N}\right) \left(\frac{1}{\sqrt{w_{ic}(gin - Ef)^2}}\right) \right\} \qquad (8)$$

Where w_{ic} is the weight of attribute i for class c.

Step 5: Obtain a comparison table from (F, A).

Step 6: Determine the score vector *S* = <*s1,s2,..sK*> for the comparison table as in Section 2.2.

Step 7: Assign the test sample to class c, where c is the class for which *sc* >*sv* for all *v* = 1, 2…*K* and c≠ *v*.

VI. EXPERIMENTAL RESULTS

The experiments are conducted for binary class problems on imbalanced medical datasets. Every classification algorithms affects with high dimensionality problem. To reduce the number of features in the imbalanced data sets a feature selection method is needed. Here we have applied a feature selection method named joint feature interaction maximization (JFIM) [25] to reduce the number of features. The subsequent section gives the details of data sets, performance evaluation measures and comparative results of proposed system with already existing algorithms.

A. *Evaluation criteria*

For experimentation, three medical imbalanced data's are collected from KEEL and UCI data set repository. All these data sets are of disease data type including Dermatology, Lymphography and Wisconsin. All these are imbalanced with uneven distribution of instances into classes. Brief sketch of input data is given in table 1.

Even though accuracy is commonly used evaluation measure to assess the efficiency of a classifier, it is not good to evaluate imbalanced datasets. The standard measures used for imbalanced data's are F-measure, AUC and G-mean. In this paper we evaluate our proposed fuzzy gravitational classifier with these evaluation measures. To compute the outcome of these measures, confusion metrics are used and it is given in Table 2.

Table 1: Datasets used for experimentation

Datasets	No. of Features (R/I/N)	No. of Samples	IR	Origin	Class(1/0)
Dermatology-6	34 (0/34/0)	358	16.9	KEEL	Positive/Negative
Lymphography-normal-	18 (0/3/15)	148	23.67	UCI	Positive/Negative

fibrosis					
Wisconsin	9 (0/9/0)	683	1.86	KEEL	Positive/ Negative

Table 2: Confusion matrix for two class problems

Predictive/Actual Results	Predicted +ve class	Predicted –ve class
Actual +ve class	TP	FP
Actual –ve class	FN	TN

True positive (TP): Actual positive class samples are correctly categorized as positive.
False positive (FP): Actual positive samples are incorrectly categorized as negative.
True negatives (TN): Actual negative samples are correctly categorized as negative.
False negatives (FN): Actual negative samples incorrectly categorized as positive.

Based on confusion metric the basic imbalanced evaluation measures could be understood in terms of following equations:
Let a, b, c, d represents the number of true positives, True negatives, false positives and False negatives respectively. The following performance evaluators are used based on the confusion metric.

$$Precision = \frac{a}{a+b} \quad (9)$$

$$Recall = \frac{a}{a+d} \quad (10)$$

$$TP_{rate} = \frac{a}{P} \quad (11)$$

$$TN_{rate} = \frac{b}{N} \quad (12)$$

$$F-measure = \frac{2*precision*recall}{precision+recall} \quad (13)$$

$$G_{mean} = \sqrt{\frac{a}{a+d} * \frac{b}{b+c}} \quad (14)$$

$$AUC = \frac{1+TP_{rate}-FP_{rate}}{2} \quad (15)$$

B. Experimental results

Our proposed method is evaluated on three medical datasets listed in Table 1. The work is compared with two well-known imbalanced algorithms via weighted nearest algorithms (WNN) and class based weighted nearest neighbor algorithms (CBWNN)[26]. Results for all mentioned evaluation measures are shown from Table 3 to Table 5 and fig 1 to 9 gives individual results. We performed our experiment by first reanking the features to find more generalized results.

Table 3: Experimental results on Dermatology Data

No. of Features	AUC			F-measure			G-mean		
	FGC	WNN	CBWNN	FGC	WNN	CBWNN	FGC	WNN	CBWNN
5	0.6823	0.6236	0.6693	.5839	.4367	.5329	.4890	.3489	.5890
10	0.7643	0.6849	0.7362	.6239	.5683	.6739	.6017	.5682	.6789
15	0.9324	0.8239	0.8694	.7867	.6783	.7290	.8683	.7892	.9012
20	0.8742	0.8692	0.932	.8356	.7360	.8190	.7932	.7290	.8367
25	0.8968	0.9021	0.8632	.8190	.7839	.8038	.7654	.7890	.8289

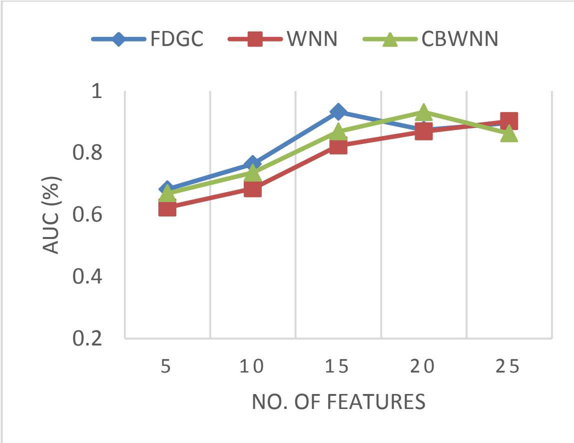
Table 4: Experimental results on Lymphography data

No. of Features	AUC			F-measure			G-mean		
	FGC	WNN	CBWNN	FGC	WNN	CBWNN	FGC	WNN	CBWNN
3	.3689	.5678	.4389	.6789	.5639	.5278	.6120	.6738	.6327
6	.6789	.6789	.7456	.8224	.7392	.7583	.8267	.7420	.7127
9	.8903	.8637	.8625	.9367	.8467	.8369	.9873	.8829	.9326
12	.9280	.9389	.8920	.9862	.8620	.8603	.9927	.8489	.8833
15	.9769	.9014	.9026	.9546	.8920	.9218	.9367	.9249	.8929

Table 5: Experimental results on Wisconsin heart disease data

No. of Features	AUC			F-measure			G-mean		
	FDGC	WNN	CBWNN	FDGC	WNN	CBWNN	FDGC	WNN	CBWNN
2	.5782	.5819	.4729	.6729	.4839	.5528	.5784	.4283	.5384
4	.7489	.8392	.6730	.7182	.6392	.6829	.6927	.6283	.6382
6	.8397	.8767	.8279	.8302	.7839	.8267	.8293	.8238	.8688
8	.9302	.8620	.8321	.9329	.9273	.8273	.9534	.9081	.9163
9	.9478	.8920	.9047	.9376	.8937	.9072	.8728	.8932	.9102

© 2018, IJCSE All Rights Reserved



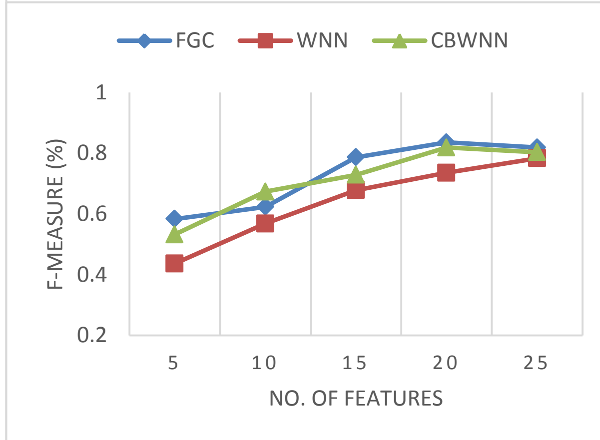
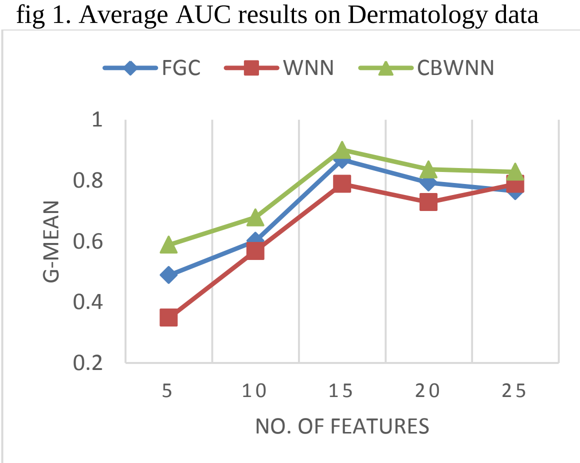


fig 1. Average AUC results on Dermatology data

fig 2. Average F-measure results on Dermatology data



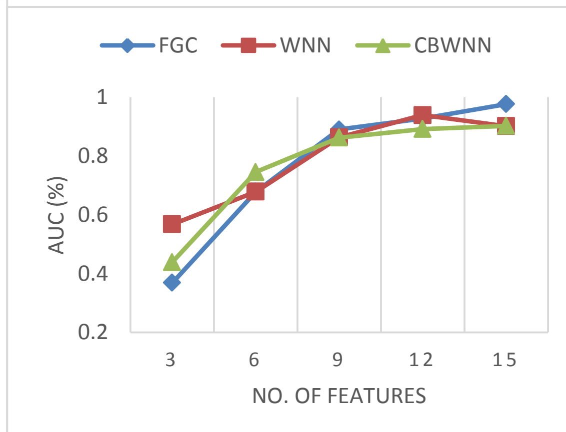
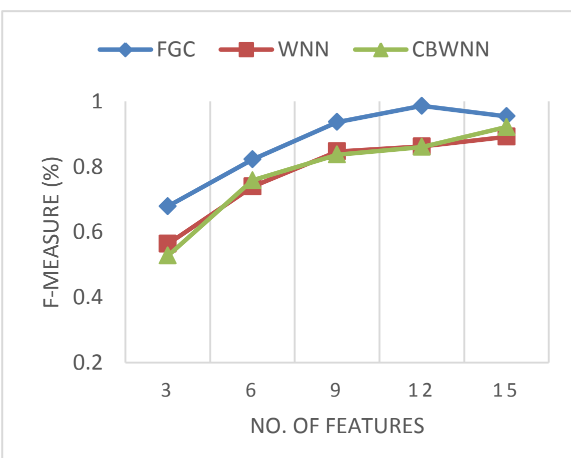


fig 3. Average G-mean results on Dermatology

fig 4. Average AUC results on lymphography



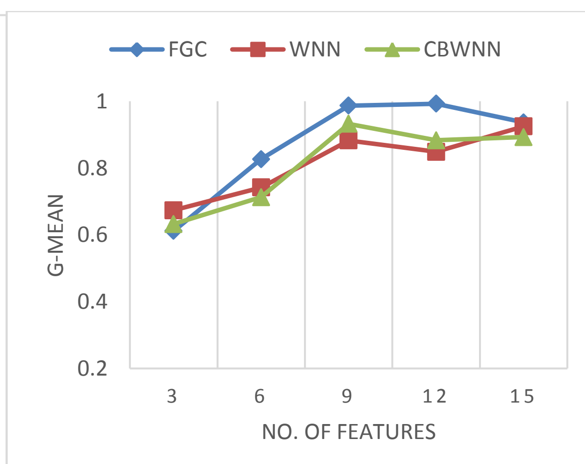


fig 5. Average F-measure results on lymphography

fig 6. Average G-mean results on lymphography

© 2018, IJCSE All Rights Reserved

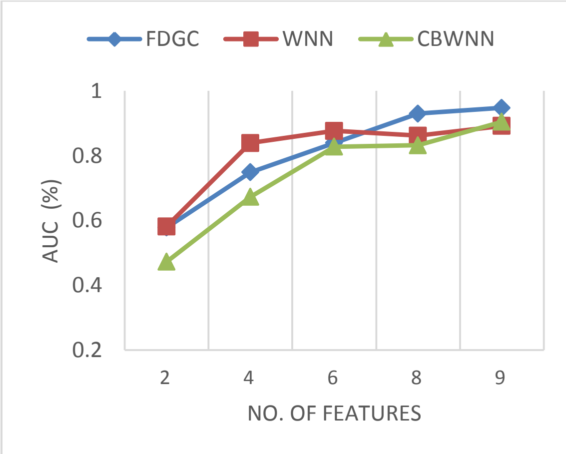


fig 7. Average AUC results on Wisconsin data

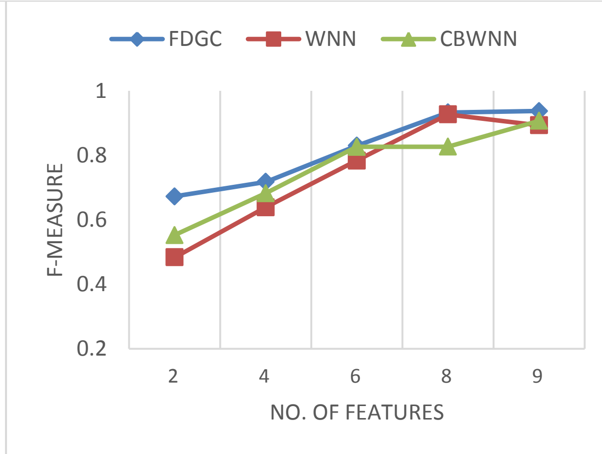


fig 8. Average F-measure results on Wisconsin data

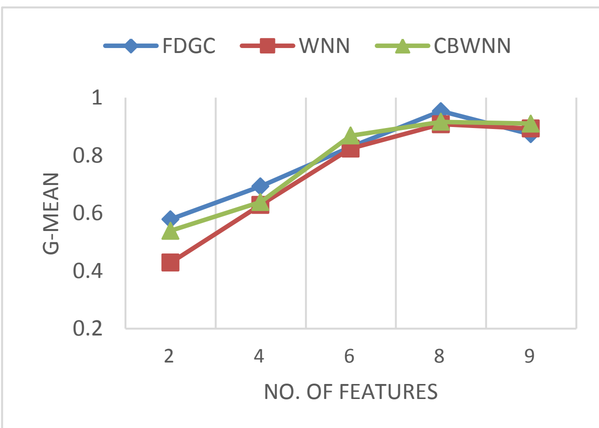


fig 9. Average G-mean results on Wisconsin data

VII. DISCUSSION

The imbalanced data considered for experimentation are Dermatology, Lymphography and Wisconsin. From this Dermatology contains a total of 34 features, Lymphography contains 18 and Wisconsin have 9 features. If we reduce the number of features, we can save execution time also quality of results can be improved. So initially we applied a filter feature selection algorithm called JFIM to get the best subset of features.

Fig. 2 illustrates the average AUC results on Dermatology dataset. It is designed by considering all the subset of selected dataset, ranges from one to 25 features. It is observed that, with 15 features it gives the best AUC results. Fig3. Shows it gives the maximum F-measure value of .83 with just 20 features. G-mean is root value of Precision and recall. From this precision evaluates the accuracy of majority class and Recall evaluates the accuracy of minority class. Dermatology gives the maximum G-mean value of .86 with 15 features.

Fig.5 to fig. 9 shows AUC, F-measure and G-mean of Lymphography and Wisconsin imbalanced data. The proposed fuzzy gravitational classifier gives better results with less number of features.

VIII. CONCLUSION

Nowadays classification of imbalanced medical data used in varied application receives more attentions in both theory and practical aspects. This paper proposes a new fuzzy gravitational classifier which gives solution to imbalancing problem. It classifies the imbalanced data based on Newton's law of gravitation. This concept is combined with fuzziness which introduces a new fuzzy membership function that can reduce the imbalance of majority class and minority class. It constructs a fuzzy equivalent relation between the unlabeled instance and gravitational function. Three benchmark datasets from UCI & KEEL are used for experimentation. This method is compared with other two imbalanced classifiers. For all datasets, FGC gives better results with less number of features. in terms of AUC, F-measure and G-mean. The proposed work has a high impact for classifying imbalanced medical data sets.

REFERENCES

[1]. R.C. Holte, L. Acker, and B.W. Porter, "Concept Learning and theProblem of Small Disjuncts" Proc. Int'l J. Conf. ArtificialIntelligence, pp. 813-818, 1989.
[2]. G.M. Weiss, "Mining with rarity: a unifying framework", SIGKDD Explor. Vol.6, pp. 7–19, 2004.

© 2018, IJCSE All Rights Reserved

[3]. V. García, R. Mollineda, J.S. Sánchez, “On the k-NN performance in a challenging scenario of imbalance and overlapping”, Pattern Anal. Appl. Vol.11, pp. 269–280, 2008.
[4]. C. Seiffert, T.M. Khoshgoftaar, J. Van Hulse, et al. ,”An empirical study of the classification performance of learners on imbalanced and noisy softwarequality data”, in: Procedings of IEEE International Conference on Information Reuse and Integration, pp. 651–658, 2007.
[5]. K. Napierala, J. Stefanowski, S. Wilk, “Learning from imbalanced data in presence of noisy and borderline examples”, in: Proceedings of 7th International Conference on Rough Sets and Current Trends in Computing (RSCTC2010), pp. 158–167, 2010.
[6]. L. Peng, H. Zhang, B. Yang, Y. Chen, M.T. Qassrawi, G. Lu, “Traffic identification using flexible neural trees”, in: Proceeding of the 18th International Workshop of QoS (IWQoS 2012), pp. 1–5, 2012.
[7]. Fawcett, T., & Provost, F. J. “Adaptive fraud detection. Data Mining and Knowledge Discovery”, Vol.1, no.3, pp.291–316, 1997.
[8]. P. Clifton, A. Damminda, and L. Vincent, “Minority Report in Fraud Detection: Classification of Skewed Data,” ACM SIGKDDExplorations Newsletter, vol. 6, no. 1, pp. 50-59, 2004.
[9]. T. Basu, C.A. Murthy, “Towards enriching the quality of k-nearest neighbour rule for document classification”, Int. J. Mach. Learn. Cybern., pp. 1–9, 2013.
[10]. Ghosh, A., Pal, N., & Das, J.,” A fuzzy rule based approach to cloud cover estimation. Remote Sensing of Environment”, vol.100, no.4, pp. 531–549, 2006.
[11]. Huang, Y. M., Hung, C. M., & Jiau, H. C.,” Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem.Nonlinear Analysis: Real World Applications”, vol.7, no.4, pp.720–747, 2006.
[12]. I. Chairi, S. Alaoui, A. Lyhyaoui,, “Intrusion detection based sample selection for imbalanced data distribution”, in: Proceeding of Second International Conference on Innovative Computing Technology (INTECH), pp. 259–264, 2012.
[13]. Xu, L., Chow, M. Y., & Taylor, L. S. (2007), ”Power distribution fault causeidentification with imbalanced data using the data mining-based fuzzyclassification E-algorithm”, IEEE Transactions on Power Systems, vol.22, no.1, pp. 164–171, 2007.
[14]. Chi, Z., Yan, H., & Pham, T. , “Fuzzy algorithms with applications to image processing and pattern recognition”. World Scientific journal off science, pp. 132-140, 1996.
[15]. Mazurowski, M., Habas, P., Zurada, J., Lo, J., Baker, J., & Tourassi, G. (2008),”Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance”, Neural Networks, vol.21, no.2, 427–436, 2008.
[16]. R. Barandela, J.S. Sánchez, V. García, E. Rangel, “Strategies for learning in class imbalance problems”, Pattern Recogn, Vol.36, pp. 849–851, 2003.
[17]. G.E.A.P.A. Batista, R.C. Prati, M.C. Monard, “A study of the behaviour of several methods for balancing machine learning training data”, SIGKDD Explor. Vol.6, pp. 20–29, 2004.
[18]. S. A. Dudani, “The distance-weighted k-nearest-neighbor rule,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. SMC-6, no. 4, pp. 325–327, 1976.
[19]. Q. Gao and Z. Wang, “Center-based nearest neighbor classifier,” Pattern Recognit., vol. 40, no. 1, pp. 346–349, 2007.
[20]. J.Wang, P. Neskovic, and L. N. Cooper, “Improving nearest neighbor rule with a simple adaptive distance measure,” Pattern Recognit. Lett., vol. 28, no. 2, pp. 207–213, 2007.
[21]. R. Paredes and E. Vidal, “A class-dependent weighted dissimilarity measure for nearest neighbor classification problems,” Pattern Recognit. Lett., vol. 21, no. 12, pp. 1027–1036, 2000.
[22]. R. Paredes and E. Vidal, “Learning weighted metrics to minimize nearest neighbor classification error,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 7, pp. 1100–1110, 2006.
[23]. L. Peng, B. Peng, Y. Chen, and A. Abraham, “Data gravitation based classification,” Inf. Sci., vol. 179, no. 6, pp. 809–819, 2009.
[24]. Sinciya P.O, J. Jeya A Celin, “Weight Optimized Gravitational Classifier for High Dimensional Numerical Data Classification “, International Journal of Pure and Applied Mathematics”, Volume 116, No. 22, pp.251-263, 2017.
[25]. Sinciya P.O, Jeyaa Celin Jesu, “JFIM: A Novel Filter Feature Selection approach Using Joint Feature Interaction Maximization”, International journal of intelligent engineering and systems, Vol.10, No.6, pp. 195-203,2017.
[26]. H. Dubey and V. Pudi. “Class based weighted k nearest neighbor over imbalanced dataset”, PAKDD 2013, pp. 305-316, 2013.

© 2018, IJCSE All Rights Reserved