PERFORMANCE IMPROVEMENT APPROACH FOR DIABETES DISEASE PREDICTION

G.Kranthikumar[1] ,K.Swathi[2]

[1]Senior Assistant Professor, Dept. of CSE, V.R Siddhartha Engineering College,[2] gkk4466@gmail.com
[2]PG Scholar, Dept. of CSE, V.R Siddhartha Engineering College,[1] swathikomma43@gmail.com

ABSTRACT

Diabetes is a major health problem all over the world. Different types of classification algorithms have applied for its treatment and diagnosis. Support vector machines and Particle swarm optimization are applied in this paper. SVM is used for classification by finding optimal hyper planes that separates one class from other class. Support vector machines are so sensitive to change their parameters. Particle swarm optimization is used for optimization technique. The proposed algorithm is implemented and organized using PIMA Indians data set which is collected from UCI repositories of machine learning database.

Keywords- Particle swarm optimization, Support vector machines, Diabetes, PIMA data set.

INTRODUCTION

Diabetes is a chronic and illness condition which affects the body's ability to use the energy found inside the food. It occurs when the level of blood glucose raises and body cannot use it properly. It doesn't produce enough insulin or it can't use the insulin or both. This disease cannot be cured but it can be managed by controlling blood glucose. It can increase the risk of cardiovascular disease.

Diabetes are of three types .they are

1. Type-1 diabetes
2. Type-2 diabetes and
3. Type-3(Gestational diabetes)

Type-1 diabetes (juvenile diabetes)

Type 1 diabetes is usually diagnosed in children and young adults. In type 1 diabetes, the body does not produce insulin. The body breaks down the sugars and starches which we eat into a simple sugar called glucose, which it uses for energy.

Symptoms

- Feeling Very thirsty.
- Feeling Very tired.
- Blurred vision caused by the lens of our eye changing shape.
- Weight loss.
- Urinating more frequently than usual, particularly at night times.
- Slow healing of cuts and grazes in our body.
- Vomiting or heavy, deep breathing can occur at a later stage if disease stage improves.

Type-2 diabetes

It is diagnosed when the pancreas does not produce enough insulin and/or the insulin does not work effectively and/or the cells of the body do not respond to insulin effectively. Usually develops in adults over the age of 45 years but is increasingly occurring in young age children, adolescents and young adults.

Causes

- It is more likely in people with a family history of type 2 diabetes or from particular ethnic backgrounds.
- Are over 45 years of age and are overweight.
- Are over 45 years of age and have high blood pressure

Symptoms

- Being more thirsty
- Passing of excessive urine
- Tired and lethargic
- Feeling hungry always
- Having cuts that heal slowly
- Skin infections and itching
- Blurred vision of eyes
- Gradually putting on more weight
- Mood swings
- Headache mostly occured
- Feeling dizzy
- Leg cramps

Type-3(Gestational diabetes)

Pregnant women who have never had diabetes before but who have high blood glucose (sugar) levels during pregnancy are said to have gestational diabetes.

Symptoms

- feel tired
- having a dry mouth
- be very thirsty
- wee a lot
- get recurring infections, such as thrush
- have blurred vision

RELATED WORK

Thirumal P. C. and Nagarajan N [1]: Jan **2015**Proposed several data mining algorithms such as Naïve Bayes, Decision trees, k nearest neighbor and SVM have been discussed and tested with Pima Indian dataset. From the experiments it is concluded that k-NN provides lower accuracy when compared to other algorithms.

Aishwarya, Anto [2]: April 2014Proposed least square support vector machine along with future subset selection, scaling and genetic algorithms. The proposed system shows a higher performance with feature subsets at an accuracy of 81.33%. They used Gaussian radial basis function is used as a kernel of LS-SVM. The robustness of the proposed system was analyzed with metrics such as a classification accuracy,

using 10-fold cross-validation and confusion matrix. The accuracy of the system for PID dataset was found to be 81.33% with GA algorithm as a feature selection method.

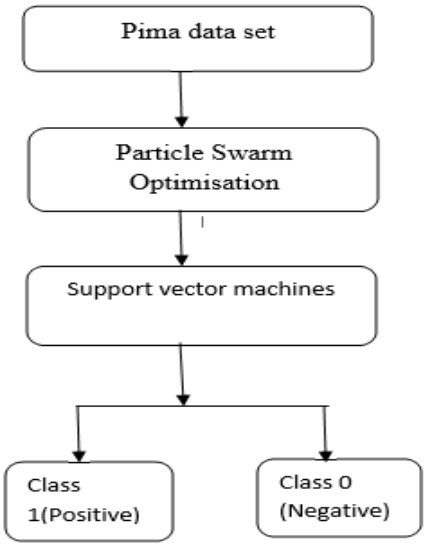
Omar s.soliman, Eman aboElhamd [3]:Feb 2014 This paper introduced a hybrid classification Algorithm for DM patients. The proposed approach Algorithm uses a modified version of PSO and LS-SVM algorithm for optimization and classification. The proposed algorithm was composed of two main phases which are Parameters Optimization and Classification. Classification had two main phases, Training phase followed by a phase of testing the algorithm. The input parameters for LS-SVM were optimized using modified version of PSO algorithm. The LS-SVM algorithm was used to classify DM patients into one of two classes (Live/Die).Modified-PSO could guarantee the robustness of the hybrid algorithm by searching for the optimal values for LS-SVM parameters.

V. Anuja Kumari, R.Chitra [4]: April 2013 In this paper, we have used datasets for diabetes disease from the machine learning laboratory at University of California, Irvine. All the patients‟ data are trained by SVM. The choice of best value of parameters for particular kernel is critical for the given amount of data SVM approach can be successfully used to detect a common disease with simple clinical measurements, without laboratory tests. In the proposed work, SVM with Radial basis function kernel is used for classification. The performance parameters are classification accuracy, sensitivity, and specificity of the SVM and RBF have found to be high so making it a good option for the classification process.

PROPOSED SYSTEM:

PIMA dataset is taken as input. The dataset consists of 10 attributes.
PSO is used for parameter optimization
Support vector machine is used for classification.
SVM consists of training and testing.
Output will be classified into classes. one class consists of patients with who have diabetes (class1) and another class consists of patients with no diabetes(class0).

Input is PIMA data set. PSO is applied for parameters optimization and then they are classified using support vector machines. Output contains two classes one is class1 which indicates patients having diabetes and another class of patients having no diabetes



PARTICLE SWARM OPTIMISATION:

Step1: In particle swarm optimization each single solution is a "bird" in the search space we call it as a "particle". All the particles have their own fitness values which are evaluated based upon the fitness function and velocities which direct the flying of particles.
Step2: The particles fly through problem space by following current particles.
Step3: PSO is initialized with group of random particles (solutions) and then searches for optima by updating generation.
Step4: In every iteration, each particle is updated by following two "best" values. The first one is best solution (fitness) which has achieved so far.
Step5: The best value is called as a local best and is denoted as lbest.
Step6: After finding the two best values, the particles will update their velocity and position.

SUPPORT VECTOR MACHINES:

Step 1: SVM algorithm performs a classification by constructing a multidimensional hyper plane which optimally discriminates the two classes by maximizing the margin between two data clusters.
Step 2: A data point is viewed as an n-dimensional vector. For example, the two variables in a dataset will create a two-dimensional space; the separating hyper plane would be a straight line (one dimensional) that divides the space into half.
Step 3: When more dimensions are involved, SVM searches for an Optimal separating hyper plane called the maximum margin separating hyper plane.
Step 4: The distance between the hyper plane and the nearest data point on each side (called support vectors) is maximized.
Step 5: Some of the data points in the two classes might fall into a "grey" area which is not easy to be separated. SVM solves this problem by

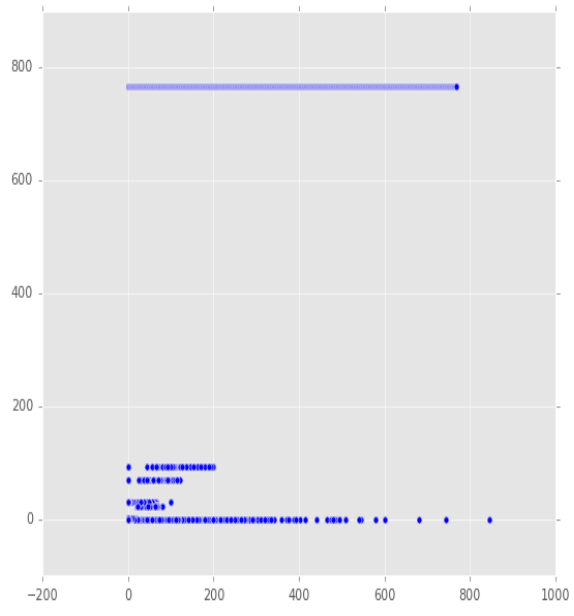
1) Allowing some of the data points to the wrong side of the hyper plane by introducing a user specified parameter C specifies the trade-off between the minimization of the misclassifications and maximization of margin;
2) Using kernel functions.

Step 6: Two key parameters for the kernels, C and gamma, need to be pre-selected to generate an optimal SVM model. Parameter C which controls the over-fitting of model by specifying the tolerance for misclassification. Parameter gamma controls degree of nonlinearity of the model.
Step 7: Test data sets were used to assess the performance of the models. Validation using the test data sets will avoid potential bias of the performance estimate due to over fitting of the model to training data sets
Step 8: Loop until stopping criteria is met, usually until reach the maximum number of iterations.

EXPERIMENTAL RESULTS:

```
            Classified as:
                     0            1
                +-------+-------+
          0     |    46 |    13 |
          1     |    15 |    26 |
                +-------+-------+

72.000 percent correct
total of 100 instances

            Classified as:
                     0            1
                +-------+-------+
          0     |   219 |    44 |
          1     |    45 |    85 |
                +-------+-------+

77.354 percent correct
total of 393 instances
>>>
```

```
SMALL DATA SET

        Classified as:
           0     1
        +-----+-----+
 0      |   51 |    8 |
 1      |   31 |   10 |
        +-----+-----+

61.000 percent correct
total of 100 instances


LARGE DATA SET

        Classified as:
           0     1
        +-----+-----+
 0      | 219 |   44 |
 1      |   64 |   66 |
        +-----+-----+

72.519 percent correct
total of 393 instances
>>>
```




```
...
['1', '6', '148', '72', '35', '0', '33.6', '0.627', '50', '1']
['2', '1', '85', '66', '29', '0', '26.6', '0.351', '31', '0']
['3', '8', '183', '64', '0', '0', '23.3', '0.672', '32', '1']
['4', '1', '89', '66', '23', '94', '28.1', '0.167', '21', '0']
['5', '0', '137', '40', '35', '168', '43.1', '2.288', '33', '1']
['6', '5', '116', '74', '0', '0', '25.6', '0.201', '30', '0']
['7', '3', '78', '50', '32', '88', '31', '0.248', '26', '1']
['8', '10', '115', '0', '0', '0', '35.3', '0.134', '29', '0']
['9', '2', '197', '70', '45', '543', '30.5', '0.158', '53', '1']
['10', '8', '125', '96', '0', '0', '0', '0.232', '54', '1']
['11', '4', '110', '92', '0', '0', '37.6', '0.191', '30', '0']
['12', '10', '168', '74', '0', '0', '38', '0.537', '34', '1']
['13', '10', '139', '80', '0', '0', '27.1', '1.441', '57', '0']
['14', '1', '189', '60', '23', '846', '30.1', '0.398', '59', '1']
['15', '5', '166', '72', '19', '175', '25.8', '0.587', '51', '1']
['16', '7', '100', '0', '0', '0', '30', '0.484', '32', '1']
['17', '0', '118', '84', '47', '230', '45.8', '0.551', '31', '1']
['18', '7', '107', '74', '0', '0', '29.6', '0.254', '31', '1']
['19', '1', '103', '30', '38', '83', '43.3', '0.183', '33', '0']
```

CONCLUSION AND FUTURE WORK

This paper includes of about classification Algorithm for DM patients. The proposed Algorithm contains PSO and SVM algorithm. The proposed algorithm mainly consists of two phases that are Parameters Optimization and Classification. Classification had two main phases, Training phase which is followed by a phase of testing the algorithm. Optimizing the parameters will minimize the classification time by avoiding making trial and error while targeting the optimal values for SVM parameters. The proposed algorithm had been implemented on Pima Indians Diabetes Data set from UCI repository of machine learning databases. As a future work, Ant Colony System (ACS) could be used as an optimization technique. Also, the other kernel functions could be applied in the classification phase.

REFERENCES

[1] Thirumal P. C. and Nagarajan N "Utilization of Data mining techniques for diagnosis of diabetes mellitus - A case study"-Jan 2015.

[2] Aishwarya, Anto "A Medical Decision Support System based on Genetic Algorithm and Least Square Support Vector Machine for Diabetes Disease Diagnosis"-April 2014.

[3] Omar s.soliman, Eman aboElhamd "Classification of Diabetes Mellitus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine"-Feb 2014.

[4] V. Anuja Kumari, R.Chitra "Classification of Diabetes Disease Using Support Vector Machine"-April 2014.