International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2017, pp. 426-429
e-ISSN:2278-621X

A COMPARATIVE STUDY OF KNN AND SVMDATA CLASSIFICATION ALGORITHMS IN CHRONIC KIDNEY DISEASE

Sayana Anil Kumar[1], Sreelakshmi K.P[2] & Vanitha T[3]

Abstract- Data Mining is becoming very popular in the field of healthcare. It is bringing new trends in the field of healthcare which will be useful for those who are working in healthcare field.Data Mining assumes an essential part to uncover new patterns in human healthcare association which thusly supportive for every one of the gatherings related with this field.The enormous measures of information created by medicinal services exchanges are excessively intricate and voluminous, making it impossible to be handled and examined by conventional methods. Data mining gives the approach and innovation to change these hills of information into valuable data for basic leadership.We have used K-NN and SVM techniques for classification in the case of chronic kidney disease(CKD).Using K-NN and SVM we have predicted chronic kidney diseases. The result shows that KNN performed better compared to SVM. K-NN has better accuracy than SVM.But it is computationally expensive.Using SVM it is easy to handle complex data points.
Keywords –SVM, KNN, Data Mining.

1. INTRODUCTION

Data mining has become important in the medical field because the demand for identifying unknown and valuable information in case of health is increasing. Nowadays data mining helps to identify meaningful information from the vast dataset. The main goal of the data mining in the medical area is better prediction through the scientific observation and experience.Data mining is a tool that is used for searching hidden values from the huge amount of data. In data mining technology healthcare data mining is the huge development area.In healthcare field/system data mining techniques are doing the active role in prediction. Some of the techniques are SVM, Decision tree, neural network and much more, here we are comparing two techniques they are SVM(Support Vector Machine) and K-NN(K-Nearest (Neighbor). Generally, SVM is the classification method and it is enlarged to receive multiple classifications.It will give better accuracy than all other available techniques. It creates hyper-planes to separate data points. K-NN is the simplest classifier.They make use of more than one nearest neighbor for classification of data points, it can be used to create early warning system in the case of chronic disease. It is used to analyze heart disease patient.Here we are taking the case of chronic kidney disease(CKD) which is also known as chronic renal disease, is an abnormal function of a kidney or in other words, we can say that it a progressive failure of renal function over a period of months or years. Chronic kidney disease is predicted using classification techniques of data mining[9]. The techniques used here is Support Vector Machine(SVM) and K-Nearest Neighbor (KNN)classifier and their performance is then evaluated based on accuracy and precision.

2. COMPARED ALGORITHM

A. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor classifier is the simplest classifier method.By making use of the previously identified data points(nearest neighbor) and classified data point the classifier will detect the unidentified data point[1].They make use of more than one nearest neighbor for classification of data points.K-NN can be used to create early warning system in case of chronic disease[2].In this case the K-NN will detect the association existing between cardiovascular disease,hypertension and risk factor of different chronic diseases[3].This classifier can also be used to analyze heart disease patient.The data is obtained from UCI. From the experiment involving both without voting or with voting K-NN classifier the comparison result was that K-NN has accuracy without voting in diagnosis of heart diseases.For diagnosing thyroid diseases Fuzzy K-NN classifier can be used. For specifying neighborhood size and fuzzy strength constraint Particle swarm optimization was used[4].

[1] Department of Software Technology, Aloysius Institute of Management and Information technology, Mangalore, Karnataka, India.
[2] Department of Software Technology, Aloysius Institute of Management and Information technology, Mangalore, Karnataka, India.
[3] Department of Software Technology, Aloysius Institute of Management and Information technology, Mangalore, Karnataka, India.

In k-NN classification, An object is differentiated by a majority vote of its neighbors, (k is a positive integer, typically small). If k = 1, to the class of single nearest neighbor the object is added .The shortest distance between two point is calculated using ,

EuclidianDistance(d)= √(x2-x1)^2+(y2-y1)^2

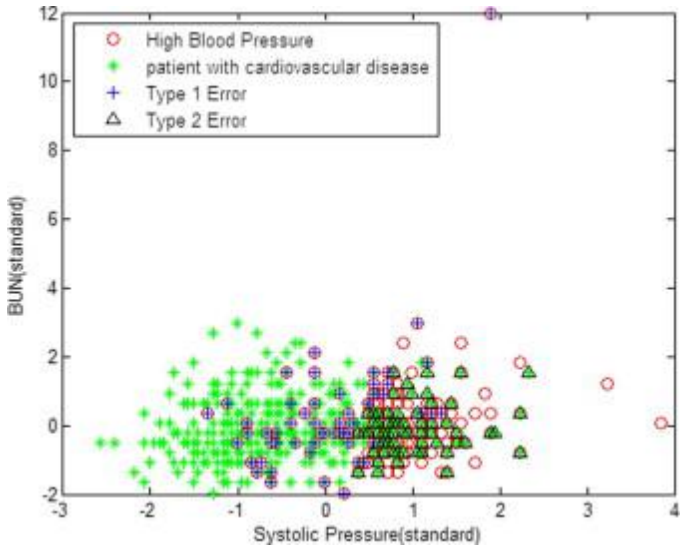

Fig.K-NN Classifier for Chronic Disease

B. Support Vector Machine (SVM):

The concept of SVM is given by Vapnik et al., is based on statistical learning theory [5-6]. This concept was originally developed for binary classification but it can be efficiently enlarged for multiclass problems also [7-8]. The SVM classifier creates a hyper plane or multi hyperplanes in high dimensional space that is useful for many efficient tasks such as classification, regression and much more. Due to this features SVM is gaining popularity and giving an efficient based performance. In original input, space SVM constructs a hyperplane in order to separate the data points. Sometimes it may be difficult to perform separation of data points in original space, so in that case, the original finite space is mapped into new higher dimensional space that will make the separation easier. The SVM concept works on the principle that using the hyperplane the data points are classified that maximize the separation between data points and the hyperplane which is constructed using support vectors.

Optimal Hyperplane for patterns: Consider the training sample $\{(xi,)\}=1$ N where xi is the input pattern for the ith instance and yi is the corresponding target output. With pattern represented by the subset yi= +1 and the pattern represented by the subset yi= -1 are linearly separable. The equation in the form of a hyperplane that does the separation is

wTx + b = 0 (1)

where x is an input vector, w is an adjustable weight vector, and b is a bias. Thus,

wTxi + b ≥ 0 for yi = +1 (2)

wTxi + b < 0 for yi = -1 (3)

For a given weight vector w and a bias b, the separation between the hyperplane defined in eq. 1 and closest data point is called the margin of separation, denoted by ρ as shown in figure 1, the geometric construction of an optimal hyperplane for a two- dimensional input space .

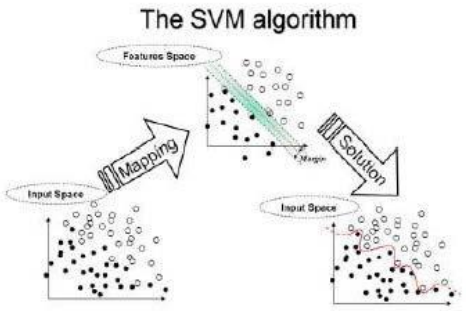

Fig.Support Vector Machine Classification

3. METHODOLOGY

In this paper We have considered two surely understood data mining algorithms. We use a data set of some patients to implement the K-NN algorithm and analyzed results. Using R programming we collect data. Then explore it. The K-NN algorithm is applied to the training data set and the results are verified on the test data set. Train a model on data. We use Euclidian method to find nearest neighbor. Then we evaluate the model performance. In the case of SVM the R interface to libsvm in package e1071, SVM(), was designed to be as natural as possible.As usual, the new data are predicted and models are shaped and both the formula interface and vector/matrix are implemented. Using R function mode can be chosen, using the depending variable's type(z): if z is a factor, then the engine will switches to classification mode otherwise it the engine presumes a novelty detection task when the z is omitted.

4. EXPERIMENT AND RESULT

We have used K-NN and SVM classification technique for comparison. The datasets are takenfrom UCI Machine learning repository to implement the K-NN algorithm and analyzed results. We have found that K-NN is having better accuracy and precision over datasets compared to SVM. The accuracy rate after comparison is shown in the Fig below

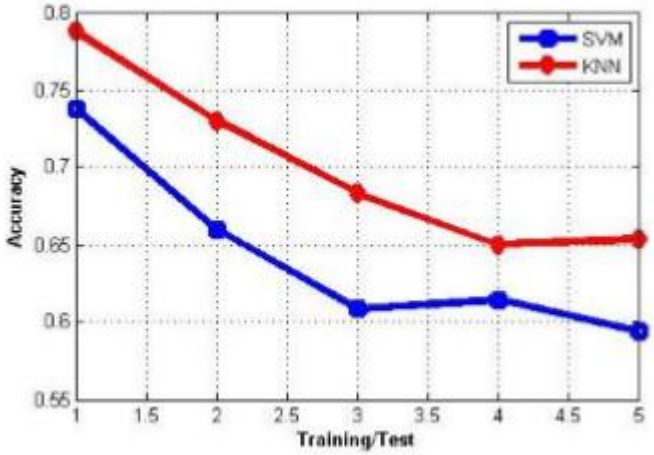

Fig. 5.1 Accuracy graph

Name of Classifier	Evaluation Parameter	
	Accuracy	**Precision**
KNN	.7873	.8567
SVM	.7373	.4995

Some of the advantage and disadvantages of both techniques are shown in the figure below:

Methods	**Advantages**	**Disadvantages**
K-NN	1. It is easy to implement. 2. Training is done in faster manner 3. Better Accuracy as compare to other classifier. 4.Better precision.	1. It requires large storage space. 2. Sensitive to noise. 3. Testing is slow.
Support Vector Machine	1. Easily handle complex nonlinear data points. 2. Over fitting problem is not as much as other methods.	1. Computationally expensive. 2. The main problem is the selection of right kernel function. For every dataset different kernel function shows different results. 3. As compare to other methods training process take more time. 4. SVM was designed to solve the

		problem of binary class. It solves the problem of multi class by breaking it into pair of two classes such as oneagainst-one and one-againstall.

5. CONCLUSION

Data mining is a standout amongst the area of research that is turning out to be progressively famous in the health organization. Through this research after comparing we found that K-NN is easy to implement, training can be done in faster way. But it require large space for storage and the testing process is very slow. KNN is having good accuracy and precision compared to SVM. SVM can easily handle complex data points .SVM is computationally expensive than K-NN. Different Kernel shows different results for every dataset. As compared to K-NN, training process take more time.

6. REFERENCES

[1] M. Silver, T. Sakara, H. C. Su, C. Herman, S. B. Dolins and M. J. O'shea, "Case study: how to apply data mining techniques in a healthcare data warehouse", Healthc. Inf. Manage, vol. 15, no. 2, (2001), pp. 155-164. "Algorithmic prediction of health-care costs", Oper. Res., vol. 56, no. 6, (2008), pp. 1382-1392.

[2] C. H. Jena, C. C. Wang, B. C. Jiangc, Y. H. Chub and M. S. Chen, "Application of classification techniques on development an early-warning systemfor chronic illnesses", Expert Systems with Applications, vol. 39, (2012), pp. 8852-8858.

[3] M. Shouman, T. Turner and R. Stocker, "Applying K-Nearest Neighbour in Diagnosing Heart Disease Patients", International Conference on Knowledge Discovery (ICKD-2012), (2012).

[4] D. Y. Liu, H. L. Chen, B. Yang, X. E. Lv, N. L. Li and J. Liu, "Design of an Enhanced Fuzzy k-nearest Neighbor Classifier Based Computer Aided Diagnostic System for Thyroid Disease", Journal of Medical System, Springer, (2012).

[5] V. Vapnik, "Statistical Learning Theory", Wiley, (1998).

[6] V. Vapnik, "The support vector method of function estimation", (1998).

[7] N. Chistianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines, and other kernel-based learning methods", Cambridge University Press, (2000).

[8] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines", Cambridge University Press, (2000).

[9] Sinha, Parul, and PoonamSinha. "Comparative study of chronic kidney disease prediction using KNN and SVM."International Journal of Engineering Research and Technology 4.12 (2015): 608-12.