IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 3, Ver. III (May – Jun. 2015), PP 21-26
www.iosrjournals.org

Support Vector Machine–Based Prediction System for a Football Match Result

Chinwe Peace Igiri
(Computer Engineering, Rivers State College of Arts and Science, Nigeria)

Abstract: *Different techniques have been used to develop result prediction systems. In particular, football match result prediction systems have been developed with techniques such as artificial neural networks, naïve Bayesian system, k-nearest neighbor algorithms (k-nn), and others. The choice of any technique depends on the application domain as well as the feature sets. The priority of a system developer or designer in most cases is to obtain a high prediction accuracy. The objective of this study is to investigate the performance of a Support Vector Machine (SVM) with respect to the prediction of football matches. Gaussian combination kernel type is used to generate 79 support vectors at 100000 iterations. 16 example football match results (data sets) were trained to predict 15 matches. The findings showed 53.3% prediction accuracy, which is relatively low. Until proven otherwise by other studies, an SVM-based system (as devised here) is not good enough in this application domain.*
Keywords: *Gaussian combination kernel, machine learning, prediction system, support vector machine*

I. Introduction

Predictive models recently have been employed to predict the weather, student performance, and stock market fluctuations. The use of machine learning and data-mining techniques to improve prediction accuracy has yielded positive results in the aforementioned fields. Consequently, it would not be out of place to apply the same techniques to football. Due to the contemporary popularity of sports, many organizations have invested a great deal to obtain better results in predicting football matches; accordingly, the prediction of game results has become an area of interest [1]. Data mining, a widely accepted method to predict and explain events, is an appropriate tool for this purpose. Various data mining techniques have been employed to predict game results in recent years, such as artificial neural networks, decision trees, Bayesian method, logistic regression, and support vector machines (SVM) and fuzzy methods.

This study seeks to study the effect of applying a system based on a support vector machine (SVM) to predict the results of football matches. The football result prediction system is a very broad area of study in computing, economics and business. For the purpose of this research, this system will be developed using data mining tools through knowledge discovery in databases (KDD). The emphasis will be on implementing the system using a SVM. Nonetheless, other related work on prediction systems will be reviewed for the purpose of completeness.

II. Related Work on Prediction Systems

A student performance prediction system has been developed [2] to identify the potential for low academic achievement in students in the beginning of an academic session to help management take informed decisions. Multi-classification techniques (i.e., algorithms such as SAMME and AdaBoost) were applied to predict student performance in an e-learning system. An MI boosting algorithm showed 80% prediction accuracy. The two boosting algorithms used in the system described in [2] were necessary for optimizing the model to get a more accurate result versus what was possible with a single classifier. A review of related literature showed that the choice of technique to a large extent depends on the parameters for the system. In some ways, data setting the yield at a relatively high prediction rate with an artificial neural network might result in low prediction accuracy when the k-nearest neighbors algorithm (k-nn) is applied [3].

Application of a general regression neural network (GRNN) and a multilayer perceptron neural network (MLPNN) was applied to predict soaked (California bearing ratio [CBR]) of remolded soil [4]. Their findings showed that GRNN was a better technique than MLPNN when applied to the soil properties used as the data set. These soil properties include gravel content, sand content, silt and clay content, liquid limit, plastic limit, soil classification, specific gravity, optimum moisture content, maximum, dry density and CBR [4].

The choice of regularization parameter C affects the performance of SVM [17]. comparative study of SVM and k-nn revealed that the k-nn classifier outperformed SVM when applied to respiratory pathogens from the lung sound database of the R.A.L.E.® Repository (rights held by PixSoft, Inc.; Winnipeg, Manitoba, Canada) [5]. In specific terms, the analysis showed 98.26% and 92.19% classification accuracies for k-nn and SVM, respectively.

In another study, a decision stump, linear regression, and SVM were used to predict stock market fluctuations [6]. A hybrid model of SVM and AdaBoost MI increased the prediction accuracy from 60% to 64% [6].

III. Theoretical framework

3.1 Machine learning

Machine learning is a branch of artificial intelligence that is concerned with building systems that require minimal human intervention in order to learn data and make accurate predictions [7]. According to Breiman [8] and Hall et al [7], in contrast to many statistical approaches, which can value inference over prediction, machine learning focuses on prediction accuracy.

Machine learning helps eliminate the static, fixed and strict approach of well-structured programming which usually provides for either poor optimization or non-efficient use of memory space and time-based factors [8].

Machine learning is composed of two phases, namely, a learning phase and a prediction phase as shown in Fig. 1. The learning phase involves the following: 1) preprocessing (normalization, reduction, data cleansing); 2) learning (supervised, unsupervised and reinforcement); 3) error analysis (precision/recall, over fitting, test/cross validation etc.); and 4) model building [9].

The prediction phase takes the output of the learning phase, which is the model to predict new data sets. The predicted data helps management or decision makers make informed decisions that are further used to build a knowledge discovery database [9].

Since this study is basically an SVM, at this point a synopsis of supervised learning is appropriate.

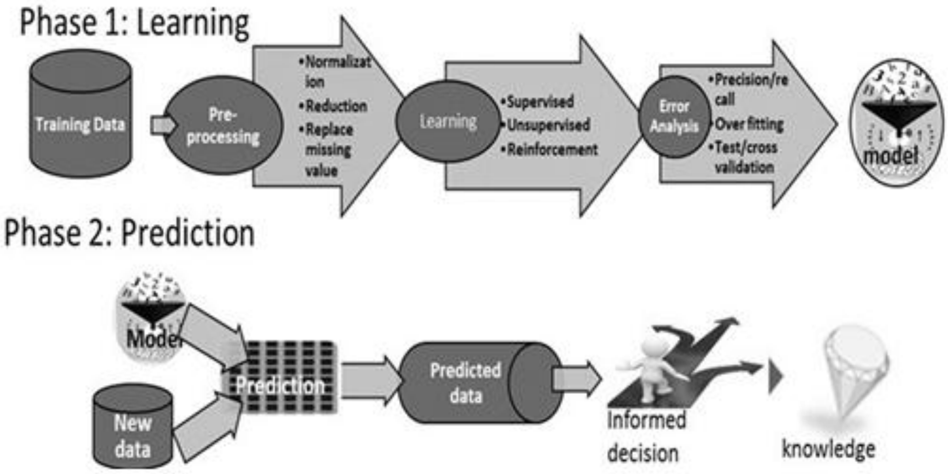


Fig.1 shows a pictorial representation of machine learning process.

3.2.1 Supervised learning

Hall et al. [7] defined supervised learning as a technique that uses labelled data to train a model. Two taxonomies of supervised learning exist, regression and classification. A regression algorithm is meant for interval labels, while a classification algorithm is for class labels [7]. Fig. 2 shows a graphical representation of a supervised machine learning classification. Here, the hyper plane classified the data sets into their respective classes, “hearts” and “faces”.

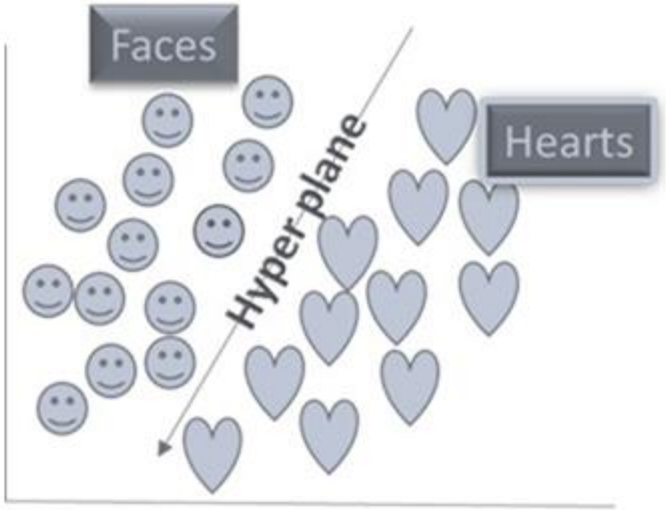


Fig.2: Graphical representation of supervised learning

Algorithms such as regression, decision tree, artificial neural network, SVM, naïve Bayesian, k-nn, Gaussian, and so forth are examples of supervised machine learning. Fig. 3 provides an illustration of supervised machine techniques.

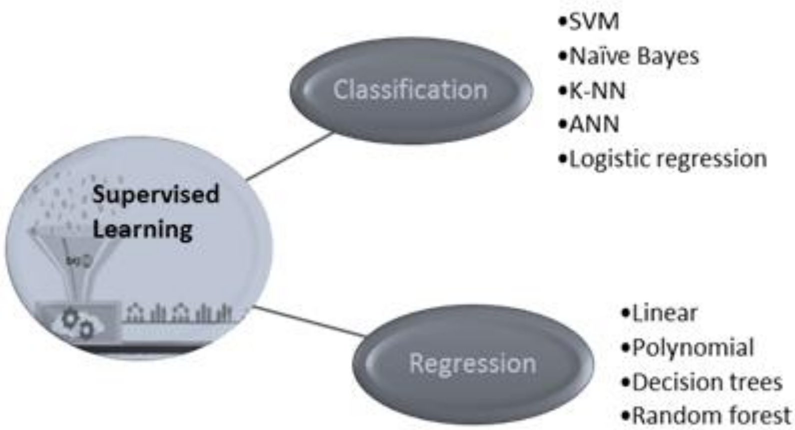


Fig.3: An illustration of supervised learning techniques

3.3 Analysis of Support Vector Machines (SVM)

Support vector machines (SVM) could provide a learning method that is used for both regression and classification, with a fast algorithm that yields good results for many learning tasks [10]. It is a non-probabilistic binary linear classifier that takes a set of input data and predicts, for each given input, which of the two possible classes comprises the input [1]. Support vectors are the training examples that comprise the support vector machine [11].

Support vector machines cannot handle nominal data, necessitating preprocessing that transforms the nominal data to numerical data. The kernel types supported by this technique are dot, radial, polynomial, neural, analysis of variance (ANOVA), Epanechnikov, Gaussian combination, multiquadric.

Dot kernel: The dot kernel is defined by the inner product of [12].

Radial kernel: The radial kernel is defined by exp(-g ||x-y||^2) where g is the gamma; it is specified by the kernel gamma parameter. The adjustable parameter gamma plays a major role in the performance of the kernel, and should be carefully tuned to the problem at hand [1].

Polynomial kernel: The polynomial kernel is defined as

$$K(x, y) = (x^\mathsf{T} y + c)^d$$

where x and y are vectors in the input space, i.e. vectors of features computed from training or test samples, and $c \geq 0$ is a free parameter trading off the influence of higher-order versus lower-order terms in the polynomial [13].

Neural kernel: The neural kernel is defined by a two-layered neural net tanh(a x*y+b), where a is alpha and b is the intercept constant. These parameters can be adjusted using the kernel a and kernel b parameters. A common value for alpha is 1/N, where N is the data dimension [1].

ANOVA kernel: The ANOVA kernel is also a radial basis function kernel, as are the Gaussian and Laplacian kernels. It is said to perform well in multidimensional regression problems [14], [15].

$$k(x,y) = \sum_{k=1}^{n} \exp(-\sigma(x^k - y^k)^2)^d$$ [14],[15]

Epanechnikov kernel: The Epanechnikov kernel is the function (3/4)(1-u2) for u between -1 and 1 and zero for u outside that range. It has two adjustable parameters, kernel sigma1 and kernel degree [1].

Gaussian combination: The Gaussian combination kernel has adjustable parameters kernel sigma1, kernel sigma2 and kernel sigma3 [1].

Multiquadric: The multiquadric kernel is also an example of a non-positive definite kernel and can be used in the same situations as the Rational Quadratic kernel [17].

$$k(x_i, x_j) = \sqrt{\|x_i - x_j\|^2 + c^2}.$$ [17]

IV. SVM Prediction System Design and Implementation

The design of the proposed system and the step-by-step implementation of the five modules are shown in Fig. 4.

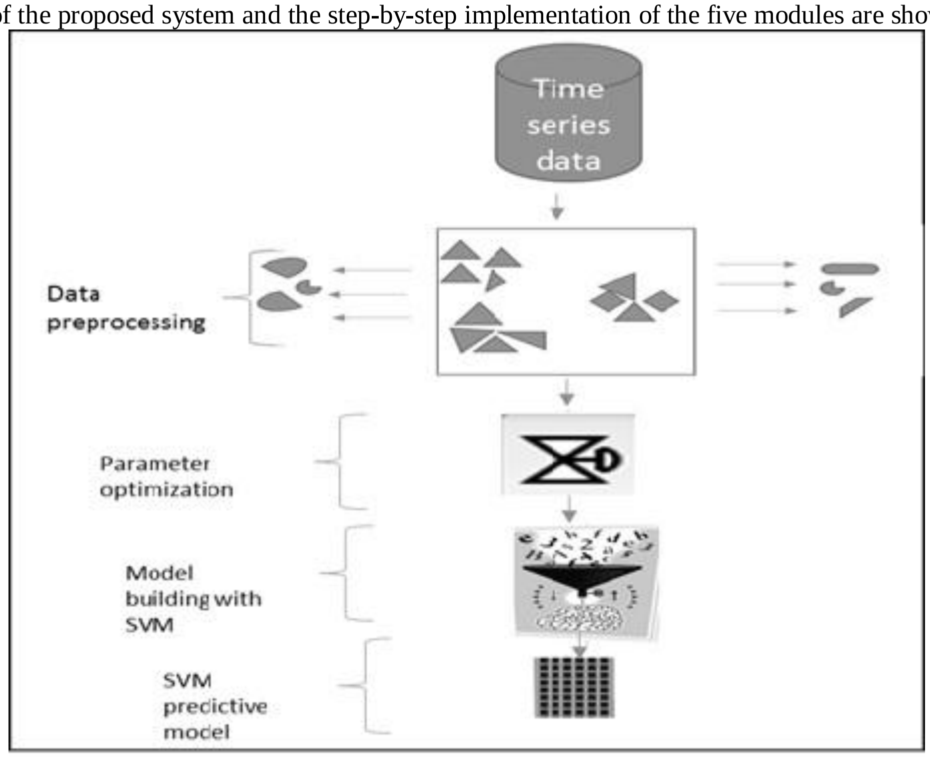


Fig4: Algorithm implementation for the SVM prediction system

Time-series data: The time-series data compose a sequence of data that is collected at regular intervals over a period of time. In this case, it is a set of data built from football match history. Players' performance and manager indices were gathered from the 2014-2015 season of the English Premier League [3].

Data pre-processing: Two data pre-processing features are introduced to the system; namely, "replace missing value" and "normalization". "Replace missing value" is used to replace missing values since SVM does not support missing value. A precise imputed missing value data-cleansing operator is used to execute this operation. It is a nested operator that always takes in data sets and returns a model. This operator calculatedly guesses missing values by learning models for each attribute (excluding the label) and applying those models to the data sets [3]. Normalization is also applied to rescale feature values to fit in a precise range. Nominal-to-numerical operators have been used to transform non–numerical values to numeric.

Parameter optimization: A Gaussian combination kernel type is used. Parameters including kernel sigma, kernel sigma2, kernel sigma3, kernel cache, constant C, convergence epsilon, and maximum iteration have been set to yield optimal prediction accuracy.

Model building using SVM: This model, as shown in Figure 4, is a non-probabilistic binary linear classifier used to train the data sets for the model. A detailed description of the parameters used for the study is in the discussion of the results.

Predictive model: This model, as shown in Figure 4, describes an SVM predictive model.

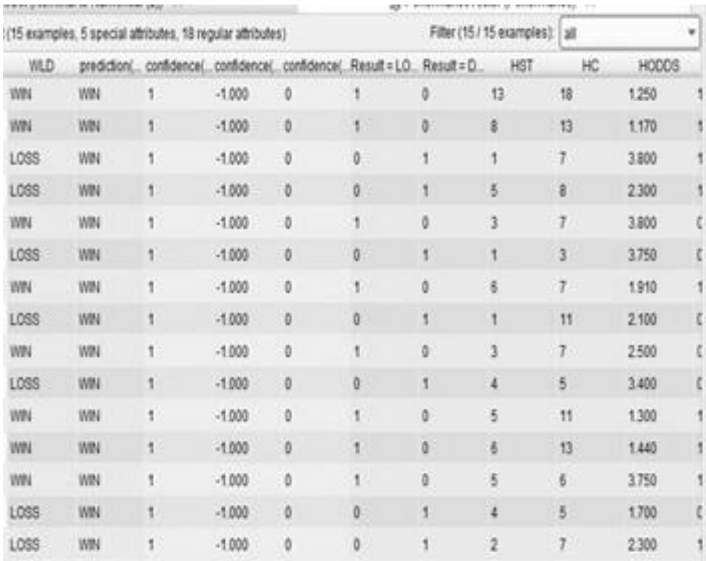
4.2. Parameters used for SVM-based Prediction System

An SVM is largely characterized by the choice of its kernel, and SVMs thus link the problems they are designed for with a large body of existing work on kernel-based methods. The following parameters are used to implement the proposed system:

Kernel type: Gaussian combination
Kernel sigma1= 1.0
Kernel sigma2= 0.0
Kernel sigma3 = 5.0
Kernel cache= 200
SVM complexity constant= 0.1
Convergence epsilon= 0.001
Maximum iterations= 100000

V. Result And Discussion

Support vector machine result
MultiModelByRegression (prediction model for label WLD)
Total number of Support Vectors: 79
Bias (offset): 1.000
w[Result = LOSS] = 3.122
w[Result = DRAW] = -3.924
w[Result = LOSS] = -2.854
w[Result = DRAW] = 3.588
w[Result = LOSS] = -0.089
w[Result = DRAW] = 0.112



(15 examples, 5 special attributes, 18 regular attributes) Filter (15 / 15 examples): all

WLD	prediction(...	confidence(...	confidence(...	confidence(...	Result = LO...	Result = D...	HST	HC	HODDS	
WIN	WIN	1	-1.000	0	1	0	13	18	1.250	1
WIN	WIN	1	-1.000	0	1	0	8	13	1.170	1
LOSS	WIN	1	-1.000	0	0	1	1	7	3.800	1
LOSS	WIN	1	-1.000	0	0	1	5	8	2.300	1
WIN	WIN	1	-1.000	0	1	0	3	7	3.800	[illegible]
LOSS	WIN	1	-1.000	0	0	1	1	3	3.750	[illegible]
WIN	WIN	1	-1.000	0	1	0	6	7	1.910	1
LOSS	WIN	1	-1.000	0	0	1	1	11	2.100	[illegible]
WIN	WIN	1	-1.000	0	1	0	3	7	2.500	[illegible]
LOSS	WIN	1	-1.000	0	0	1	4	5	3.400	[illegible]
WIN	WIN	1	-1.000	0	1	0	5	11	1.300	1
WIN	WIN	1	-1.000	0	1	0	6	13	1.440	1
WIN	WIN	1	-1.000	0	1	0	5	6	3.750	1
LOSS	WIN	1	-1.000	0	0	1	4	5	1.700	[illegible]
LOSS	WIN	1	-1.000	0	0	1	2	7	2.300	1

Fig.5: Screen shoot of SVM based prediction system

The time-series football data was imported into the Rapid Miner studio. Two preprocessing algorithms (“replace missing value” and “nominal to numerical”) were used to transform the data, since SVM does not work with missing values and nominal data. The learner (SVM) was adjusted to the parameters specified in the parameter listing in order to increase the model performance. The kernel type for the proposed system is Gaussian combination; the specific values for the kernels and for other parameters (SVM complexity constant C, convergence epsilon, maximum iteration) appear directly before this section. Seventy-nine support vectors were generating for training 15 feature sets used to build the model. Sixteen new data sets were used to test and validate the model. The execution time was 15 seconds, resulting in 53.3% prediction accuracy. It was also observed that when 60 training sets were used, the execution time was more than one hour, showing that SVM does not support large data sets (in contrast to artificial neural networks) [3].

VI. Conclusion and further work

This study investigated the performance of an SVM multimodel by regression for prediction of the results of football matches using the English Premier League as data sets. A Gaussian combination kernel type of SVM was used to analyze the football feature set. A total of 38 attributes were used for each match. Prediction accuracy was 53.3% when 16 data sets were trained to predict unknown matches result for 15 matches; 8 out of the 15 were predicted correctly, while 7 were offset.

The findings shows that SVM is not an appropriate technique for feature sets used for this analysis, since the same feature set yielded 85% prediction accuracy using an artificial neural network [3].

The limitation to this study is low prediction accuracy. Further research can be carried out on how to improve prediction accuracy using SVM. Other kernels, including dot, ANOVA, Epanechnikov, multiquadric, polynomial, radial, and so forth could be investigated to verify their performance in terms of prediction accuracy.

References

[1]. F. Akhtar and C. Hahne. Rapid Miner 5 Operator Reference. Rapid-I GmbH, 2012, Retrieved February 13, 2015 from: http://rapidminer.com/wpcontent/uploads/2013/10/RapidMiner_OperatorReference_en.pdf .

[2]. A. Malaise, A. Malibari, M. Alkhozae, "Student Performance Prediction System Using Multi-agent Technique," International Journal of Data Mining and Knowledge Process (IJDKP), vol. 4(5), 2014, 1–20.

[3]. C.P. Igiri, E.O. Nwachukwu, "An Improved Prediction System for a Football Match Result," IOSR Journal of Engineering (IOSRJEN), 4(12), 2014, 12–20.

[4]. S. A. Kumar, J.P. Kuma, J. Rajeev, "Application of Machine Learning to Predict Soaked CBR of Remolded Soils," International Journal of Engineering Research and Technology (IJERT), 4(6), 2013, 3019–3024.

[5]. R. Palaniappan, K. Sundaraj, and S. Sundaraj, "A Comparative Study of SVM and K-NN Machine Algorithms for Diagnosis of Respiratory Pathologies using Pulmonary Acoustic Signal," BMCC Bioformatics, 15, 2014, 223.

[6]. H.S. Vatsal, "Foundation of Machine Learning," Spring, New York University, 2017.

[7]. P. Hall, J. Dean, I.K. Kabul, J. Silva, "An Overview of Machine Learning with SAS Enterprise Miner," SAS Institute Inc, 2014. Retrieved May 6, 2015 from http://support.sas.com/resources/papers/proceedings14/SAS313-2014.pdf.

[8]. L. Breiman, "Random Forests," Machine Learning, vol. 45, 2001, 5–32.

[9]. C.P. Igiri, O.U. Anyama, A.I. Silas, I. Sam, "A Comparative Analysis of K-NN and ANN Techniques in Machine Learning," International Journal of Engineering Research and Technology (IJERT), vol. 4(3), 2015, 420–425.

[10]. A. Boz, "Large Scale Machine Learning using NVIDIA CUDA," retrieved May 7, 2015, from http://www.codeproject.com/Articles/336147/Large-Scale-Machine-Learning-using-NVIDIA-CUDA, 2012. Technical blog.

[11]. S. Rüping (2000), mySVM-Manual, University of Dortmund, Lehrstuhl Informatik 8, http://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/).

[12]. B. Scholkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New Support Vector Algorithms," Neural Computation 12, °c 2000 Massachusetts Institute of Technology, (2000), 1207–1245.

[13]. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Section 16.5. Support Vector Machines". Numerical Recipes: The Art of Scientific Computing (3rd ed.), 2007, New York: Cambridge University Press. ISBN 978-0-521-88068-8.

[14]. R. Khardon , L. DeYoung, 2013. http://www.cs.tufts.edu/~roni/Teaching/CLT/LN/lecture18.pdf

[15]. 1171–1220.

[16]. C. Souza, "Kernel Functions for Machine Learning Applications," retrieved May 7, 2015 from http://crsouza.com/2010/03/kernel-functions-for-machine-learning-applications/#anova.

[17]. Q. S. Zhang, "A Multi-Label Learning Based Kernel Automatic Recommendation Method for Support Vector Machine," DOI: 10.1371/journal.pone.0120455, 2015.