Computing, Information Systems, Development Informatics & Allied Research Journal
Vol. 9 No. 2, June, 2018 - www.cisdijournal.org

Seeking Intelligent Convergence for Asymptotic Stability Features of the Prey/Predator Retarded Equation Model Using Supervised Models

Ojugo, A.A. & Yoro, R.E
Department of Mathematics/Computer Science
Federal University of Petroleum Resources Effurun
Effurun, Delta State, Nigeria.
E-mails: ojugo.arnold@fupre.edu.ng, arnoldojugo@yahoo.com, arnoldojugo@gmail.com,
otakore.oghenevwede@fupre.edu.ng

ABSTRACT

Optimization tasks seeks to resolve complex task by exploring knowledge as well as exploits historic data to yield new paradigm, whose outcomes is the optimal solution of the underlying probability of possible features. The challenges in modeling time-varying phenomenon using retarded equation, has since been of great interest to mathematicians as it helps to address the non-instantaneous reaction of state parameters. Our study modifies Volterra prey/predator model via time-lag functions of state parameters that change. The contraction and continuity feats on Banach space are utilized to establish the uniqueness of integral solution of critical point. We investigate asymptotic stability feats using quadratic matrix equation and symmetric linear matrix inequality test to seek model-based solution. Result shows that for asymptotical stable model, if the recruitment rate of prey is kept higher than recruitment rate of the predator at all time in the system – it yields stronger condition for stability and sustainability of system; Compared to the stability result of the equivalent ordinary differential equation model of system. Based on computational solution, we selected supervised models to find convergence. And study shows KNN outperforms QDA and LDA to yield an approximate solution that converges closer to an analytic solution. This can be easily extended to solve a wide range of other problems.

Keywords: Volterra Prey/Predator, Retarded Equation, QDA, LDA, KNN, quadratic function

CISDI Journal Reference Format
Ojugo, A.A. & Yoro, R.E. (2018): Seeking Intelligent Convergence for Asymptotic Stability Features of the Prey/Predator Retarded Equation Model Using Supervised Models. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 8 No 2. Pp 13-26
Available online at www.cisdijournal.org

1. INTRODUCTION

The increased interest in the use of retarded equations in the modeling of dynamic systems – can be attributed to the fact that traditional point-wise modeling assumption lapse of differential equations is addressed by incorporating response time $h > 0$ in the reacting state parameters. For more on the analysis of retarded models equation, see the following: Brauer and Castillo (2001), Wodarz and Nowark (2002), Gaetano and Arino (2004), Igobi et al (2007); Hinow et al (2007) and Abderrazzak et al (2009). In this study, the prey-predator model by Volterra (1928) is linearized and modified by introducing the time-lag functions $f(t-h)$ into the model equations, and thus, account for the past history of the state parameters (Igobi et al, 2012).

The compactness, contraction and continuity properties of the functional on the Banach space are utilized to establish the uniqueness of the integral solution of the critical point of the modified model equation. The asymptotic stability properties of the critical solution are checked using the quadratic matrix equation and symmetric linear matrix inequality test (Igobi, 2009) on asymptotic stability analysis of solutions of retarded systems. This test ensures that the coefficient matrix of the model equation satisfies the quadratic matrix equation and the symmetric linear matrix inequality (LMI), for some choice positive definite symmetric matrices. Also, the eigen-values of the characteristic equation of coefficient matrix should have negative real roots for some chosen values of the rate of the state parameters. These established properties enhanced the predictive capability of the model.

A. Modeling Dynamic and Chaotic Phenomenon

Chaotic and dynamic phenomena and events abound in all spheres of our daily endeavors. These must be resolved via the creation of systems and models that mimic such behaviors. Thus, nonlinear systems are modeled as mathematical models using equations to resolve its nonlinearity, dynamism and complex feats. There exists various methods in use to resolve quadratic or differential equations – though, optimal solution for such systems and models are still a herculean challenge (Fletcher and Reeves, 1964; Forsythe, 1986 and Friedlander et al, 1999). Studies have shown to solve a first order differential equation, we can use unsupervised neural network model as a factor of speed (Ghalambaz et al, 2011). Lagaris et al (1998) went further to represent a new method to solve first-order linear ordinary and partial equations using ANN. Khan et al (2009) provided a hybrid ANNPSO model to solve Wessinger equation (though the model could not satisfy initial boundary conditions). It was improved by Ghalambaz et al (2011); while Khan et al (2009) used the fuzzy trained neural network in evaluating the well-known Wessinger's quadratic function as a constraints satisfaction problem. Their results show the model converged closer to data points in the range of their analytical solution. Allenotor (2016) adopted memetic algorithm and SA to evaluate quadratic function. A trial solution of the quadratic equation is also written as sum of two parts: (a) to satisfy initial condition for unconstrained optimization using DFP and hybrids as separate methods, and (b) apply DFP as a pre-processor with adjustable parameters of for ANN-TLRN hybrid.

B. Mathematical Optimization

Mathematical optimization is yielding of best element from set of available alternatives, maximizing or minimizing a real function that symmetrically choose inputs from an allowed set of variables, to compute the function's output via the objective function in a given domain. It exploits domain knowledge that is expressed as mathematical model to yield a method that is both tolerant to noise, accepts partial truth, uncertainty and ambiguities at its input – while, exhibiting great robustness, flexibility, and continuous adaptation in its bid to resolves constraints within the system (Ojugo et al, 2013).

It yields output feats with uncontrolled parameters as input (not explicitly present from outset); But, modeled therein via boundary values in domain space confined to real parameters and derives via experience, ability to recognize behavioral feats from observed data, and can suggest optimal solution of high quality, void of over-fitting, irrespective of modification made to model via other approximations with multiple agents. These constantly affect the quality of the solution. Also, many of such heuristics are combined to create hybrid(s) that seeks to explore the structural differences in the statistical-method(s) used, and help resolve the implications of the conflict constrained on the model by such a multi-agent populated system – as agents can often create their own behavioral rules based on historic dataset (Ojugo et al, 2014; 2015).

C. Notations

A and B are nxn matrices such that the matrix L = (A+B).P is a positive symmetric matrix, whereas the matrix Q = I is an n×n identity matrix. E^n is the n-dimensional Euclidean space, with $\|\bullet\|$ as the Euclidean vector norm. $B_H([t-h,t],E^n)$ is the Banach space of continuously differentiable functions f, G on $[t-h,t)$. $\varphi(s)$ is a continuously function with norm in $B_H([t-h,t],E^n)$ defined as $\|\varphi(s)\| = \sup_{t-h\le s<t} |\varphi(s)|$ such that $\varphi(0)=0$. W is a symmetric matrix with diagonal elements $0 \le a_{ii} \le 1$ for each w_i defines i^{th} row of W, $0 \le w_{ii} \le 1$, and $\zeta = \{x, y \in E^n; w_i x - (1-w_i)y\} \in W$ is the convex set segment of W.

D. The Model Formulation

Volterra (1928) considers x(t) being the population at time t of some animal species known as the prey, and y(t) being the population of predator species which lives on these preys. It is assumed that x(t) increases at a rate proportional to x(t) if preys were left alone, and x(t) = ηx(t), η>0. In the presence of a predator, which depends on finding prey for food, the growth of the prey is then hindered by how much the predator finds the prey, thus: $x(t) = \eta x(t) - lx(t)z(t), \quad \eta, l > 0$. The predator is assumed to depend wholly on the prey for its food.

Thus, in the absence of the prey $y(t) = -kx(t) \quad k > 0$.
Volterra (1928) presented the differential equation model as,

$$\begin{aligned} \dot{x}(t) &= \eta x(t) - lx(t)z(t) \\ \dot{z}(t) &= lx(t)z(t) - kz(t), \\ x(t_0) &= x_0,, \quad z(t_0) = z_0, \end{aligned} \qquad (1.0)$$

Linearizing system (1.0) at critical points $(\alpha, \Gamma) = (0,0), (\frac{k}{l}, \frac{\eta}{l})$ using the definition of the deviation of state parameters from their equilibrium points as: $\frac{d}{dt}(\alpha + \delta x(t)) = \delta \dot{x}(t)$ and $\frac{d}{dt}(\Gamma + \delta z(t)) = \delta \dot{z}(t)$, then

$$\delta \dot{x}(t) = \eta(\alpha + \delta x(t)) - l[(\alpha + \delta x(t))(\Gamma + \delta z(t))]$$

$$\delta \dot{z}(t) = l[(\alpha + \delta x(t))(\Gamma + \delta z(t))] - \eta(\Gamma + \delta x(t)).$$

And linearized system (1.0) yields:

$$\begin{aligned} \delta \dot{x}(t) &= \eta\, \delta x(t) - l\alpha\delta z(t) - l\Gamma\delta x(t) \\ \delta \dot{z}(t) &= l\alpha\delta z(t) + l\Gamma\delta x(t) - k\,\delta z(t), \\ x(t_0) &= x_0 \ \ and \ \ z(t_0) = z_0. \end{aligned} \qquad (1.1)$$

For critical point $(\alpha, \Gamma) = (0,0)$, the system (1.1) yields:

$$\begin{aligned} \delta \dot{x}(t) &= \eta\, \delta x(t) \\ \delta \dot{z}(t) &= -k\, \delta z(t), \\ x(t_0) &= x_0 \ \ and \ \ z(t_0) = z_0, \end{aligned} \qquad (1.1a)$$

Its trivial solution ($x = 0, \quad z = 0$) is seen not to be stable. Thus, we consider the linearization at the critical points given by $(\alpha, \Gamma) = (\frac{k}{l}, \frac{\eta}{l})$, and system (1.1) yields system (1.1b):

$$\begin{aligned} \delta \dot{x}(t) &= -k\delta z(t) \\ \delta \dot{z}(t) &= \eta\delta x(t), \\ x(t_0) &= x_0 \ \ and \ \ z(t_0) = z_0, \end{aligned} \qquad (1.1b)$$

which has characteristic root $\lambda = \pm\, i\sqrt{\eta k}$, so the critical point is stable but not asymptotically stable.

E. Model Modification

Assume that the prey and the predator take average reacting time $h > 0$ to respond to changes in the system; that is after the predator have killed many of the prey, the predator begin to die out after a time-lag $t - h,\ h > 0$ due to food shortage. And in turn gives the prey pool a chance to recover within the same time-lag. Thus the retarded equation model from (1.1) is:

$$\begin{aligned} \delta \dot{x}(t) &= \eta\, \delta x(t-h) - l\alpha\delta z(t) - l\Gamma\delta x(t) \\ \delta \dot{z}(t) &= l\alpha\delta z(t) + l\Gamma\delta x(t) - k\,\delta z(t-h), \\ x(t_0) &= x_0 \ \ and \ \ z(t_0) = z_0. \end{aligned} \qquad (1.2)$$

F. Uniqueness of Solution of Modified Model Equation

Consider system (1.2) in the form (1.3):

$$\dot{V}(t) = AV(t) + BV(t-h), \qquad V(t_0) = \varphi(s), \quad t-h \le s < t, \tag{1.3}$$

where, $\dot{V}(t) = \begin{pmatrix} \dot{\delta x}(t) \\ \dot{\delta z}(t) \end{pmatrix}$, $A = \begin{pmatrix} -l\Gamma & -l\alpha \\ l\Gamma & l\alpha \end{pmatrix}$, $B = \begin{pmatrix} \eta & 0 \\ 0 & -k \end{pmatrix}$, $V(t) = \begin{pmatrix} \delta x(t) \\ \delta z(t) \end{pmatrix}$

$V(t-h) = \begin{pmatrix} \delta x(t-h) \\ \delta z(t-h) \end{pmatrix}$ *and* $\varphi(t) = \begin{pmatrix} x(t_0) \\ z(t_0) \end{pmatrix}$,

So, that the integro-differential equation of (1.3) is defined as:

$$\frac{d}{dt}\left(\varphi(t) + B\int_{t-h}^{t} \varphi(s)ds\right) = (A+B)\varphi(t) \cdot \tag{1.4}$$

2. MODIFIED RETARDED EQUATION: MATERIAL & METHOD

A. Definition

If $\varphi(s) \epsilon B_H$ for $t-h \le s < t$, then a function $V(s,\varphi(s))$ is an integral solution of the critical point (α, Γ) of system (1.3) on $[t-h, t)$ if the following conditions hold:

a. We have that $V(s,\varphi(s))$ is continuous on [t-h, t]

b. $V(s,\varphi(s))$ = V(t), for t – h $\le$ s < t

c. $\int_{l-k}^{t} \varphi(s)ds \in B_H$ and

d. From the system (1.4) we have that $V(s,\varphi(s)) = \varphi(0)L + B\int_{t-h}^{t}\left(\int_0^s \varphi(\varpi)d\varpi\right)ds - L\int_0^t \varphi(s)ds$, for $t-h \le s < t$,

The integral solution of (1.3) at critical point $\Omega = (\alpha, \Gamma)$ is: $V_\Omega(s,\varphi(s)) = \varphi(0)L + B\int_{t-h}^{t}\left(\int_0^s f_\Omega(\varpi,\varphi(\varpi))d\varpi\right)ds - L\int_0^t G_\Omega(s,\varphi(s))ds$. (1.5)

By definition (1.0), if hypothesis hold for $V(s,\varphi(s)) \in B_H$;

a. Functions $f_\Omega(\varpi,\varphi(\varpi)),\ G_\Omega(s,\varphi(s))$ forms compacts set $D \subset B_H$, where $D = \left\{\left(V(x(t),z(t),s) : \left\|V(s,\varphi(s)) - \varphi(s)\right\| < \varepsilon,\ s \in [t-nh,t),\ n \ge 1\right)\right\}$. For all $\{f_i\}_{i=1}^{\infty}, \{G_i\}_{i=1}^{\infty}$, convergent subsequences are selected and $\{f_i\}_{i=1}^{\infty} \to \varphi$ *and* $\{G_i\}_{i=1}^{\infty} \to \varphi$ in $D \subset B_H$.

b. Also $f_\Omega : [t-h,t) \times D \to E^n$ and $G_\Omega : [t-h,t) \times D \to E^n$ are continuously differentiable, and satisfies contraction condition on B_H. Then, there exist constants k_0, m_0 for $0 < k_0 < 1$ and $0 < m_0 < 1$ such that $\left\|f_\Omega(\varpi,\varphi(\varpi)) - f_\Omega(\varpi,\theta(\varpi))\right\| \le m_0\left\|\varphi - \theta\right\|$ and $\left\|G_\Omega(s,\varphi(s)) - G_\Omega(s,\theta(s))\right\| \le k_0\left\|\varphi - \theta\right\|$, *for* $\varphi, \theta \in B_H$

B. Theorem 1

Let B and L be stable matrix, and assume hypothesis a, b are satisfied, then for a given $\varphi(s) \in B_H$, there exists a unique solution of the critical point $\Omega = (\alpha, \Gamma)$ of system (1.3) defined as $V_\Omega(\ s, \varphi(s)\ ) \in B_H$.

Proof:

Let $V(s, \varphi(s))$ be a continuous differentiable function on $[t-h, t)$, with vector norm defined as $\|\bullet\|$ on B_H. Since $V(s, \varphi(s))$ satisfies conditions i – ii, then equation (1.5) defined the integral solution of the critical point of system (1.3). By the hypothesis of compactness of $f_\Omega(\varpi, \varphi(\varpi)), G_\Omega(s, \varphi(s))$ in D for $t_o < t_1 < \ldots < t_n$, there exists convergence subsequence solution given by the system $\{V_\Omega(s, \varphi_i(s))\}_{i=1}^n \in D$ which converges in D. Furthermore, using hypothesis a, b and defining $\Lambda_{\max}\lambda$, $V_{\max}\lambda$, as the maximum characteristic roots of matrices B and L respectively, and $e^{t\Lambda_{\max\lambda}}$, $e^{tv_{\max\lambda}}$ as the corresponding eigen-vectors, then we have that:

$$\|V_\Omega(s,\varphi(s))\| = \left\| \left(B\int_{t-h}^{t} \left(\int_0^s f_\Omega(\varpi,\varphi(\varpi))d\varpi \right) ds + L\int_0^t G_\Omega(s,\varphi(s))ds \right) \right\|$$

and

$$\|V_\Omega(s,\theta(s))\| = \left\| \left(B\int_{t-h}^{t} \left(\int_0^s f_\Omega(\varpi,\theta(\varpi))d\varpi \right) ds + L\int_0^t G_\Omega(s,\theta(s))ds \right) \right\|$$

so we now have that:

$$\|V_\Omega(s,\varphi(s)) - V_\Omega(s,\theta(s))\|$$

$$\leq \|B\| \left(\left\| \int_{t-h}^{t} \left(\int_0^s f_\Omega(\varpi,\varphi(\varpi))d\varpi \right) ds - \int_{t-h}^{t} \left(\int_0^s f_\Omega(\varpi,\theta(\varpi))d\varpi \right) ds \right\| \right) + \|L\| \left(\left\| \int_0^t G_\Omega(s,\varphi(s))ds - \int_0^t G_\Omega(s,\theta(s))ds \right\| \right)$$

$$= \|B\| \left(\left\| \int_{t-h}^{t} f_\Omega(\varphi(s)ds - \int_{t-h}^{t} f_\Omega(s,\theta(s)ds \right\| \right) + \|L\| k_0 \| v_\Omega(t,\varphi(t)) - v_\Omega(t,\varphi_\Omega(t)) \|$$

$$= m_0 e^{t\Lambda_{\max\lambda}} \|w_\Omega(s,\varphi(s) - w_\Omega(s,\theta(s)\|_{t-h\leq s<t} + k_0 e^{tv_{\max\lambda}} \|v_\Omega(t,\varphi(t)) - v_\Omega(t,\varphi_\Omega(t))\|$$

$$\leq m_0 e^{t\Lambda_{\max\lambda}} \sup_{\varphi\in B_H} |w_\Omega(s,\varphi(s) - w_\Omega(s,\theta(s)|_{t-h\leq s<t} + k_0 e^{tv_{\max\lambda}} \sup_{\varphi\in B_H} |v_\Omega(t,\varphi(t)) - v_\Omega(t,\varphi_\Omega(t))|$$

Since $0 < m_0 < 1$, and $0 < k_0 < 1$ then

$$m_0 e^{t\Lambda_{\max\lambda}} \sup_{\varphi\in B_H} |w_\Omega(s,\varphi(s) - w_\Omega(s,\theta(s)|_{t-h\leq s<t} + k_0 e^{tv_{\max\lambda}} \sup_{\varphi\in B_H} |v_\Omega(t,\varphi(t)) - v_\Omega(t,\varphi_\Omega(t))| < 1$$

Then $V_\Omega(s, \varphi(s))$ is a strict contraction in D and the fixed point of $V_\Omega(s, \varphi_i(s))$ for $t - h \leq s < t$ which is defined as $V_\Omega(t)$ is the unique integral solution of the critical point of system (1.3).

C. Conditions for asymptotic stability

Consider system (1.3) with the coefficient matrix as

$$A = \begin{pmatrix} -l\Gamma & -l\alpha \\ l\Gamma & l\alpha \end{pmatrix}, \quad B = \begin{pmatrix} \eta & 0 \\ 0 & -k \end{pmatrix},$$

Han (2001) defined the characteristic polynomial equation of the coefficient matrix of (1.3) as defined in (1.5) at the critical point $\alpha = \frac{k}{l}$ *and* $\Gamma = \frac{\eta}{l}$ (which is dependent on the time-lag $h > 0$) as

$$f(e^{-h},\lambda)=\begin{vmatrix}-\eta+\eta e^{-h}-\lambda & -\eta \\ k & k-ke^{-h}-\lambda\end{vmatrix}=0,$$

$$f(e^{-h},\lambda)=\lambda^2+\lambda((\eta-k)-(\eta-k)e^{-h})+\eta k(2e^{-h}-e^{-2h})=0 \quad (1.6).$$

Igobi et al (2009) stated the necessary and sufficient conditions for system (1.3) to assume asymptotic stability as:

1. All real roots of the characteristic polynomial equation of (1.6) must have negative real value
2. For any positive definite symmetric matrix P, the fundamental matrix $L=A+Be^{-h}$ must satisfy the quadratic matrix equation $L^TP+PL=-Q$ (1.7)

   For Q being an identity matrix and $L=\begin{bmatrix}-\eta+\eta e^{-h} & -\eta \\ k & k-ke^{-h}\end{bmatrix}$
3. Given matrices Q, P, L as stated in (ii) above and matrix B as in (1.3), then the exists a symmetric matrix W which ensures that the symmetric linear matrix inequality **(LMI)**

$$\begin{bmatrix}\sum_{11} & \sum_{12} & \sum_{13} \\ \sum_{21} & \sum_{22} & \sum_{23} \\ \sum_{31} & \sum_{32} & \sum_{33}\end{bmatrix}\leq 0, \quad (1.8)$$

for $\sum_{11}=\left|-Q+BW^TP^T\right|$

$\sum_{22}=|WB|$ and

$\sum_{33}=\left|PBM^T\right| and \sum_{ij}=0, \; for \; i\neq j$ is satisfied (see proof in Igobi et al, 2009).

D. Asymptotic Stability Analysis

The asymptotic stability conditions of section (1.4) are satisfied if the following holds:

1. The rate of recruitments in the system must be such that $\eta>k$ (that is, the rate of recruitment of the prey η must at all times be greater than the rate of recruitment of the predator k), then the characteristic polynomial equation of (1.6) will have negative real roots and this satisfies condition (i) for system (1.3) to be asymptotically stable.
2. Resolving the matrix equation (1.7) into

$$\begin{pmatrix}2m_{11} & 2m_{21} & 0 \\ m_{12} & (m_{11}+m_{22}) & m_{21} \\ 0 & 2m_{12} & 2m_{22}\end{pmatrix}\begin{pmatrix}p_{11}\\p_{12}\\p_{22}\end{pmatrix}=\begin{pmatrix}-1\\0\\-1\end{pmatrix},$$

For m_{11}, m_{12}, m_{21}, m_{22} being the elements of matrix **L**, and substituting the values to have

$$\begin{pmatrix}2(-\eta+\eta e^{-h}) & 2\eta & 0 \\ -k & -\eta+\eta e^{-h}+k-ke^{-h} & \eta \\ 0 & -2k & 2(k-ke^{-h})\end{pmatrix}\begin{pmatrix}p_{11}\\p_{12}\\p_{22}\end{pmatrix}=\begin{pmatrix}-1\\0\\-1\end{pmatrix}.$$

Thus, elements of the positive symmetric matrix **P** becomes:

$$p_{11} = -\left[(k^2 - 4\eta k) - (k^2 - \eta k)e^{-h}\right],$$

$$p_{12} = \left[(2k^2 - \eta k) + (-2\eta^2 - \eta k)e^{-h}\right],$$

$$p_{22} = -\left[\eta k + (-\eta^2 - 2\eta - 4\eta k)e^{-h} + (2\eta^2 + 2\eta k)e^{-2h}\right]$$

$$p_{00} = 2\left[3\eta^2 k + (2\eta k^2 - 5\eta^2 k)e^{-h} + (3\eta^2 k - 3\eta k^2)e^{-2h} + (\eta^2 k - \eta k^2)e^{-3h}\right],$$

Thus, the positive definite symmetric matrix P which satisfies the quadratic matrix equation (1.7) for condition (ii) to be fulfilled is

$$P = \frac{1}{p_{00}}\begin{pmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{pmatrix} \qquad (1.9)$$

For $p_{00}, p_{11}, p_{12}, p_{22}$ as defined above and $\eta, k > 0$.

c. the choice of the symmetric matrix W that satisfied the symmetric linear matrix inequality (1.8) for condition (iii) to be fulfilled is

$$W = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

$0 \le a_{ii} \le 1$ and $w_i\ (for\ 0 < w_i \le 1)$ defines i^{th} row of W

Intelligent Proposed Model

We seek to compare resolution of above quadratic function via the supervised models below using unsupervised model (HMM as benchmark) to measure comparative performances.

Linear Discriminant Analysis (LDA)

LDA is a simple and effective supervised classification method with wide range of applications. Its basic theory is to classify compounds (rules) dividing n-dimensional descriptor space into two regions separated by a hyper-plane that is defined by linear discriminant function. Discriminant analysis generally transforms classification tasks into functions that partitions data into classes; Thus, reducing the problem to an identification of a function. The focus of discriminant analysis is to determine this **functional** form (assumed to be linear) and estimate its coefficients. LDA was first introduced in 1936 by Ronald Aylmer Fisher and his LDA function works by finding the mean of a set of attributes for each class, and using the mean of these means as boundary. The function achieves this by projecting attribute points onto the vector that maximally separates their class means and minimizes their within-class variance as expressed in Eq. 17 as follows:

$$LDA = X'S^{-1}(X_2 - X_1) - \frac{1}{2}(X_2 + X_1)'S^{-1}(X_2 - X_1) > c \qquad (17)$$

where X is vector of the observed values, X_i (i = 1, 2…) is the mean of values for each group, S is sample covariance matrix of all variables, and c is cost function. If the misclassification cost of each group is considered equal, then c = 0. A member is classified into one group if the result of the equation is greater than c (or = 0), and into the other if it less than c (or = 0). A result that equals c (set to 0) indicates such a sample cannot be classified into **either** class, based on the features used by the analysis. LDA function distinguishes between two classes – if a data set has more than two classes, the process must be broken down into multiple two-class problems. The LDA function is found for each class versus all samples that were not of that class (one-versus-all). Final class membership for each sample is determined by LDA function that produced the highest value and is optimal when variables are normally distributed with equal covariance matrices. In this case, the LDA function is in same direction as Bayes optimal classifier (Billings and Lee, 2002), and it performs well on moderate sample sizes in comparison to more complex method (Ghiassi and Burnley, 2010). Its mathematical function is simple and requires nothing more complicated than matrix arithmetic. The assumption of linearity in the class boundary, however, limits the scope of application for linear discriminant analysis. Real-world data frequently cannot be separated by linear boundary.

When boundaries are nonlinear, the performance of the linear discriminant may be inferior to other classification methods. Thus, to curb this – we adopt a decimal encoding of the data to give us a semblance of linear, continuous boundaries.

Quadratic Discriminant Analysis (QDA)
QDA is another distance-based classifier by Smith (1947), which is very similar to and more of an extension of LDA. Both discriminant functions assume that values of each attribute in each class are normally distributed, however, the discriminant score between each sample and each class is calculated using the sample variance –covariance matrix of each class separately rather than the overall pooled matrix and so is a method that takes into account the different variance of each class. While, LDA assumes that the covariance matrices of the groups are equal; QDA makes no assumption. When the covariance matrices are not equal, the boundary between the classes will be a hyper-conic and in theory, the use of quadratic discriminant analysis will result in better discrimination and classification rates. However, due to the increased number of additional parameters to be estimated, it is possible that the classification by QDA is worse than that of linear discriminant analysis (Malhotra et al. 1999). The QDA is found by evaluating the Eq. 18:

$$QDA = X'(S_1^{-1} - S_2^{-1})X + 2\,(Y_2' S_1^{-1} - Y_1' S_2^{-1})X - [Y_2' S_1^{-1} Y_2 - Y_1' S_2^{-1} Y_1 + Ln(\frac{|S_2|}{|S_1|})] > c \quad (18)$$

The same conditions apply to the nature of c as well as the classification, in the case that the result is equal to c or zero. As with LDA, the QDA distinguishes between two classes. For multiple class data sets, this was handled the same as for linear discriminant analysis. Size of differences in variances determines how much better QDA performs better than LDA. For large variance differences, QDA excels when compared to LDA. Additionally, of the two, only QDA can be used when population means are equal. QDA is more broadly applicable than the LDA; But, less resilient in non-optimal conditions. The quadratic discriminant can behave worse than the linear discriminant for small sample sizes. Additionally, data that is not normally distributed results in a poorer performance by the quadratic discriminant, when compared to the linear discriminant. Marks and Dunn (1974) found the performance of the quadratic discriminant function to be more sensitive to the dimensions of the data than the linear discriminant, improving as the number of attributes increases to a certain optimal number, then rapidly declining. Linear and nonlinear discriminant functions are the most widely used classification methods. This broad acceptance is due to their ease of use and the wide availability of tools. Both, however, assume the form of the class boundary is known and fits a specific shape. This shape is assumed to be smooth and described by a known function. These assumptions may fail in many cases. In order to perform classification for a wider range of real-world data, a method must be able to describe boundaries of unknown, and possibly discontinuous, shapes.

E. K-Nearest Neighbourhood (KNN)
The K-nearest neighbour (KNN) model is a well-known supervised learning algorithm for pattern recognition that first introduced by Fix and Hodges in 1951, and is still one of the most popular nonparametric models for classification problems (Fix and Hodges 1951; 1952). K-nearest neighbour assumes that observations, which are close together, are likely to have the same classification. The probability that a point x belongs to a class can be estimated by the proportion of training points in a specified neighbourhood of x that belong to that class. The point may either be classified by majority vote or by a similarity degree sum of the specified number (k) of nearest points. In majority voting, the number of points in the neighbourhood belonging to each class is counted, and the class to which the highest proportion of points belongs is the most likely classification of x.

The similarity degree sum calculates a similarity score for each class based on the K-nearest points and classifies x into the class with the highest similarity score. Its lower sensitivity to outliers allow majority voting to be commonly used other than the similarity degree sum (Chaovalitwongse, 2007). We use majority voting for the data sets to determine which points belongs to neighbourhood so that distances from x to all points in the training set must be calculated. Any distance function that specifies which of two points is closer to the sample point could be employed (Fix and Hodges, 1951). The most common distance metric used in K-nearest neighbour is the Euclidean distance (Viaene, 2002).

The Euclidean distance between each test point ft and training set point fs, each with n attributes, is calculated via Eq. 19:

$$d = [(f_{t1} - f_{s1})^2 + (f_{t2} - f_{s2})^2 \dots + (f_{tn} - f_{sn})^2]^{\frac{1}{2}} \quad (19)$$

In general the following steps are performed for the K-nearest neighbour model (Yildiz et al., 2008): (a) chosen of k value, (b) distance calculation, (c) distance sort in ascending order, (d) finding k class values, (e) finding dominant class.

A challenge in using Knn is to determine optimal size of k, which acts as smoothing parameter. A small k is not sufficient to accurately estimate the population proportions around the test point. A larger k will result in less variance in probability estimates (but for risk of introducing more bias). K should be large enough to minimize probability of a non-Bayes decision, and small enough that all points included, gives an accurate estimate of the true class. Enas and Choi (1986) found optimal value of k to depend on sample size and covariance structures in each population and on the proportions for each population in the total sample. For cases where differences in covariance matrices and difference between sample proportions are both small or both large, it is found that optimal k is $N^{3/8}$ (N is number of samples in the training set). If and when there is a large difference between covariance matrices, and a small difference between sample proportions (or vice-versa), the optimal value k is determined by $N^{2/8}$ (Enas and Choi, 1986).

This model presents several merits (Berrueta et al., 2007) in that: (a) its mathematical simplicity does not prevent it from achieving classification results as good as (or better than) other more complex pattern recognition techniques, (b) it is free of statistical assumptions, (c) its effectiveness does not depend on the space distribution of classes, and (d) when the boundaries between classes are not hyper-linear or hyper-conic, K-nearest neighbour performs better than LDA.

But, Enas and Choi (1986) found LDA performs slightly better than K-nearest neighbour if the population covariance matrices are equal, a condition that suggests linear boundary. As differences in covariance matrices increases, k-nearest neighbour performs increasingly better than LDA and QDA function. However, despite these merits, the demerits of k-nearest neighbour model include that it does not work well if large differences are present in samples in each class. K-nearest neighbour provides poor data about the structure of its classes, and relative importance of variables in classification. Also, it does not allow graphical representation of the results, and in case of large number of samples, computation become excessively slow. In addition, K-nearest model requires more memory and processing requirements than other methods. All prototypes in training set must be stored in memory and used to calculate Euclidean distance from every test sample. The computational complexity grows exponentially as the number of prototypes increases (Muezzinoglu and Zurada, 2006).

3. RESULT FINDINGS AND DISCUSSION

To measure their effectiveness and classification accuracy, we adopt the misclassification rate of each model as well as its corresponding improvement percentages of the proposed model in comparison with those of other classification models for the dataset (in both training and test data) summarized in Table 2. The equations for its improvement percentage when trained as below:

$$\text{Improvement Percentage} = \frac{\text{MR(A)} - \text{MR(B)}}{\text{MR(A)}} \times 100 \quad (20)$$

Table 1: Improvement Percentage

Model	Improvement %	
	Training Data	Testing Data
LDA	41.16	45.83
KNN	48.29	64.2
QDA	41.79	46.09

Results show from Tables 1 shows our supervised models in LDA, QDA and KNN showed an improvement rate of 45.8%, 64.2% and 46.1% respectively. Also, it is observed that though the KNN scores were sensitive to relative magnitude of different attributes, all attributes are scaled by their z-scores before using KNN model. This is in tandem with Antal et al (2003) and results are supported by Khasei et al (2012), Perez and Marwala (2011), Mandic and Chambers (2001), Meade and Fernandez (1994), Ojugo et al (2015a) and Allenotor (2016). The number of data-points in quadratic function has great influence on model performance – such that, if the data points are too small, most supervised models may not achieve its accuracy. If it is too many, it can result also in overtraining as well as over-parameterization.

The rationale for the choice of models includes:

a. Unsupervised Model have great capability in learning to approximate function – making them flexible and mostly, universal estimator cum approximator. Their adaptive feat is a huge merit in modeling dynamic, changing states. However, the problem is that of encoding, fitness function selection, parameterization and training/retraining model to suit the purpose for which is seeks to understudy the underlying probability feats. But, when these are taken care of – unsupervised models perform excellently well.
b. Supervised Model such as LDA, QDA and SVM (support vector machines) have their demerits as in section II.

4. CONCLUSION AND RECOMMENDATIONS

The prey/predator model equation was modified with a time-lag function, to account for time response of the reacting state parameters. The formulation and proof (mathematically) of result on uniqueness of the integral solution of the critical point of the modified model equation was achieved by the utilization of the compactness, contraction and continuity properties of the functional on the Banach space. The modified model equations show desirable asymptotic stability properties under a certain condition at the critical point (such as keeping the recruitment rate of the prey higher than the recruitment rate of the predator). This is not the case with the equivalent ordinary differential system (1.1), whose solutions have shown that the system is only be stable, but not asymptotically stable at the critical point (the later being a stronger condition for sustainability of the system).

Study consists of 4-phases: (a) train models using quadratic function, (b) delete outliers in data by resolving the quadratic function and finding underlying probability on convergence ability, (c) calculate membership probability of output points in each class, and (d) assign output to appropriate class by largest probability. We adopt three (3) known intelligent and statistical classification models: LDA, QDA and KNN. It is observed that KNN outperforms both LDA and QDA model respectively.

The unsupervised models do not assume the shape of the partition, unlike the linear and quadratic discriminant analysis. In contrast to K-nearest neighbour model, the proposed model does not require storage of training data. Once the model has been trained, it performs much faster than K-nearest neighbour does, because it does not need to iterate through individual training samples. The proposed model does not require experimentation and final selection of a kernel function and a penalty parameter as is required by the support vector machines. Our proposed model solely relies on a training process in order to identify the final classifier model. Finally, the unsupervised models does not need large amount of data in order to yield accurate results.

REFERENCES

[1] Armijo, S., (1966).Minimization function having Lipschitz continuous partial derivatives, Pacific Journal of Mathematics, 16, pp 1-3.
[2] Barzilar, J and Borwein, J.W.,(1988).Two point step size gradient methods, IMA Journal of Numerical Analysis, pp141-148.
[3] Berardi, V. and Zhang, G. P., "The effect of misclassification costs on neural network classifiers", Decision Sciences, Vol. 30, pp. 659– 68, 1999.
[4] Billings, S. and Lee, K., "Nonlinear Fisher discriminant analysis using a minimum squared error cost function and the orthogonal least squares algorithm", Neural Networks, Vol. 15, pp. 262– 270, 2002.
[5] Broyden, C.G., (1967). Quasi newton method and their application to function approximation, Mathematical Computation, 21, pp 368-381.
[6] Calisir D. and Dogantekin, E., "An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector Machine Classifier", Expert Systems with Applications, Vol. 38, pp. 8311– 8315, 2011.
[7] Chaovalitwongse, W., (2007). On the time series k-nearest neighbor classification of abnormal brain activity, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol. 37
[8] Dai, Y.H.,(2001). Alternate step gradient method, Report AMSS-2001-041, Academy of Mathematics and Systems Science, pp 32-56.
[9] Dai, Y.H and Yuan, Y.,(2003). Alternate minimization gradient method, IMA Journal of Numerical Analysis, 23, pp 377-393.
[10] Dai, Y.H and Yuan, Y.,(1999). A nonlinear conjugate gradient method with strong global convergence property, SIAM Journal of Optimization, 10, pp 177-182.
[11] Dai, Y.H., Yuan, J.Y and Yuan, Y.,(2002). Modified two-point stepsize gradient methods for unconstrained optimization, Computational Optimization and Application, 22, pp 103-109.
[12] Daniel, J.W.,(1967). Conjugate gradient method for linear/nonlinear operator equations, SIAM J. Numeric Analysis, 4, pp10 – 26.
[13] Dennis, J.E and Moore, J.J.,(1974). A characterization of superlinear convergence and its application to Quasi Newton methods, Mathematical Computation, 28, pp 549-560.
[14] Dennis, J.E and Moore, J.J.,(1977). Quasi Newton method: motivation and theory, SIAM Rev, 19, pp 46-89.
[15] Denton, J.W and Hung, M.S.,(1996). A comparison of nonlinear optimization methods for supervised learning in multilayer feedforward networks, European J. of Operation Research, 93, pp 358-368.
[16] Dixon, L.C.W.,(1972). Variable metric algorithms necessary and sufficient conditions for identical behaviour on non-quadratic functions, J. of Optimization Theory and Application, 10, pp34-40.
[17] Enas, G. and Choi, S., (1986). Choice of the smoothing parameter and efficiency of k-nearest neighbor, Computers and Mathematics with Applications, Vol. 12, pp. 235– 244.
[18] Fix, E and Hodges, J., (1951). Discriminatory analysis – nonparametric discrimination: Consistency properties, Project No. 21-49-004, Report No. 4, Contract No. AF 41(128)-31, USAF School of Aviation, Randolph Field, Texas, 1951.
[19] Fletcher, R.,(1987) Practical methods of optimization, John Wiley and Sons, Chichester.
[20] Fletcher, R and Reeves, C.,(1964). Function minimization by conjugate gradients, Computational Journal, 7, pp 149-154.
[21] Forsythe, G.E.,(1986). On asymptotic directions of the s-dimension optimum gradient method, Numerische Mathematik, 11, pp57-76.
[22] Friedlander, A., Martinez, J.M., Molina, B and Raydan, M.,(1999). Gradient method with retards and generalization, SIAM J. of Numerical analysis, 36, pp275-289.
[23] Ghalambaz, M., Noghrehabadi, A.R., Behrang, M.A., Assareh, E., Ghanbarzadeh, A and Hedayat, N.,(2011). Hybrid gravitational search neural network method to solve the well known Wessinger's equation, World Academy of Science and Technology, 49, pp 803 – 807.
[24] Giles, D and Draeseke, R., (2001). Economic Model Based on Pattern recognition via fuzzy C-Means clustering algorithm, Economics Dept Working Paper, EWP0101, University of Victoria, Canada.
[25] Gottlieb, D and Orszag, S.A.,(1977). Numerical analysis of spectral methods: theory and applications, CBMS-NSF Regional Conference Series in Applied Mathematics, 26.
[26] Griewank, A and Toini, P.H.,(1982). Local convergence analysis of partitioned Quasi Newton updates, Numerical Mathematics, 39, pp 429-448.

[27] Hager, W and Zhang, H.,(2003). A new conjugate gradient method wit guaranteed descent and an efficient line search, SIAM J. of Optimization, pp305-333.

[28] Inan, G and Elif, D., (2005). Adaptive neuro-fuzzy inference system for classification of EEG using wavelet coefficient, retrieved from http://www.rorylewis.com/PDF/04

[29] Khan, J., Zahoor, R and Qureshi, I.R,(2009). Swarm intelligence for problem of non-linear ordinary differential equations and its application to well known Wessinger's equation. European J. Sci. Research, 34(4), pp514-525.

[30] Lagris, I.E., Likas, A and Fotiadis, D.I.,(1998). Artificial neural networks for solving ordinary and partial differential equation, IEEE Transactions Neural Network, 9(5), pp987–1000.

[31] Lee, H and Kang, I.S.,(1990). Neural algorithms for solving differential equations, Journal of Computational Physics, 91, pp110–131.

[32] Liu, Y and Storey, C.,(1991). Efficient generalized conjugate gradient algorithms, J. of Optimization Theory Application, 69, pp129-137.

[33] Malek, A and Beidokhti, R.S.,(2006). Numerical solution for high order differential equations using hybrid neural network - Optimization method, Applied Mathematics and Computation, 183, pp260-271.

[34] Mandic, D. and Chambers, J.,(2001). Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability, John Wiley and Sons: New York.

[35] Meade, A.J and Fernandez, A.A.,(1994). The numerical solution of linear ordinary differential equations by feedforward neural networks, Mathematical and Computer Modeling, 19(12), pp1–25.

[36] Nash, J and Sutcliffe, J., (1970). River flow forecasting with conceptual models, J. of Hydro. Sci., 10, pp.282–290.

[37] Ojugo, A.A, (2012a).Artificial neural networks gravitational search model for rainfall runoff simulation and modeling, unpublished PhD, Computer Sci. Department, Ebonyi State University Abakiliki, Nigeria.

[38] Ojugo, A.A., (2015): A comparative stochastic model solution on the convergence problem of quadratic functions, Unpublished Thesis, Mathematics and Computer Science Department, Federal University of Petroleum Resources Effurun

[39] Ojugo, A.A., Eboka, A.O., Okonta, E.O., Yoro, R.E and Aghware, F.O.,(2012b). Genetic algorithm rule-based intrusion detection system, Journal of Emerging Trends in Computing and Information Systems, 3(8), pp 1182 - 1194.

[40] Ojugo, A., and Yoro, R.,(2013a). Computational intelligence in stochastic solution for Toroidal Queen, Progress in Intel Comp. App., 2(1), doi: 10.4156/pica.vol2.issue1.4, pp46–56.

[41] Ojugo, A.A., Emudianughe, J., Yoro, R.E., Okonta, E.O and Eboka, A.,(2013b). Hybrid artificial neural network gravitational search algorithm for rainfall runoff modeling in Hydrology, Progress Intel. Comp. App., 2(1), doi: 10.4156/pica.vol2.issue1.2, pp 22 – 33.

[42] Ojugo, A.A., Allenotor, D., Oyemade, D.A., Yoro, R.E and Anujeonye, C.N., (2015a): Immunization Model for Ebola Virus in Rural Sierra-Leone, African Journal of Computing & ICTs. Vol 8, No. 1, Pp 1-10.

[43] Ojugo, A.A., Allenotor, D., Oyemade, D.A., Longe, O.B and Anujeonye, C.N., (2015b): Comparative stochastic study for credit-card fraud detection models, African Journal of Computing & ICTs. Vol 8, No. 1, Issue 1. Pp 15-24.

[44] Perez, M and Marwala, T., (2011). Stochastic optimization approaches for solving Sudoku, IEEE Transaction on Evol. Comp., pp.256–279.

[45] Pham, D and Karaboga, D.,(1999). Training Elman and Jordan networks for system identification using genetic algorithms, Artificial Intelligence in Engineering, 13, pp.107–117.

[46] Pham, D., Koc, E., Ghanbarzadeh, A and Otri, S.,(2006). Optimization of weights of multi-layered perceptrons using bees algorithm. Proceedings of Intelligent Manufacturing Systems. Sakarya University, Department of Industrial Engineering, pp38–46.

[47] Pham, D.T and Liu, X.,(1995). Artificial Neural Networks for Identification, Prediction and Control, Springer Verlag, London.

[48] Plumb, A.P., Rowe, R.C., York, P and Brown, M.,(2005). Optimization of the predictive ability of artificial neural network models, European J. of Pharmaceutical Sciences, 25, pp395-405.

[49] Polak, E and Ribiere, G.,(1969). Note sur la convergence de directions conjugees, Rev. Francaise Informat Researche Operationelle, 3(16), pp35-43.

[50] Polyak, B.T.,(1969). The conjugate gradient method in extreme problems, USSR Computation Mathematics: Mathematic Physics, 9, pp94-112.

[51] Powell, M.J.,(1971). On the convergence of the variable algorithm D, Winston Mathematics Application, pp21-38.

[52] Powell, M.J.,(1976). Some global convergence properties of a variable metric algorithm for minimization without exact line searches, In Cottle, R.W and Lemke, C.E (eds.), Nonlinear programming, SIAM Proceedings of AMS, 9, pp53-72.
[53] Raydan, M.,(1993). On Barzilai and Borwein choice of stepsize for the gradient method, IMA Journal Numeric Analysis, 13, p321-326.
[54] Reynolds, R.,(1994). An introduction to cultural algorithms, IEEE Transaction on Evolutionary Programming, pp.131-139.
[55] Ritter, K.,(1979). Local and superlinear convergence of a class of variable method, Computing, 23, pp287-297
[56] Stachurski, A.,(1981). Superlinear convergence of a class of variable method, Mathematical Programming, 14, pp178-205.
[57] Ursem, R., Krink, T., Jensen, M.and Michalewicz, Z.,(2002). Analysis and modeling of controls in dynamic systems. IEEE Transaction on Evolutionary Computing, 6(4), pp.378-38
[58] Abderrazzak, B., Hassane, B. andAbdelhakim, M. (2009). Neutral Volterra Integrodifferential Equations with Infinite Delay, Int. Journal of Math. Analysis, Vol. 3, no.4,` 187-196
[59] Brauer, F., and Castillo-Chavez, C. (2001). Mathematical Model of the Population Biology and Epidemiology, Springer, New York.
[60] De Gaetano, A., Arino, O. (2004). Modeling of the intra venous Glucose Tolerence Test; a global study for single-distributed-delay model, Discrete and Dynam. System Journal, B4, 407-417.
[61] Hinow, P. Wang, S. E., Arteaga, C. L. and Genn, F. (2007). A Mathematical Model Separating Quantitatively the Cytostatic and Cytotoxic Effects of HER2 Tyrosine Kinase Inhibitor, javascript return, (10), 25-38.
[62] Igobi, D. K., Egwurube, M. O. & Adesola , W. A. (2007). On the Existence and Uniqueness, and Analysis of Asymptotic Stability Properties of Retarded Systems, International Journal of Natural and Applied Sciences (IJNAS), Vol. 1, no 1, 116-122.
[63] Igobi, D. K., Eguurube, M. O. and Odekunle, M. R. (2009). Asymptotic Stability Results for Retarded Differential Systems, Global Journal of Mathematical Sciences (8) no.1, 23-29
[64] Volterra, V. (1928). Variation and fluctuation of the number of individuals in animal species living together, Animal Ecology.
[65] Wodarz, D and Nowark, M. A (2002). Mathematical models of HIV pathogenesis and treatment bioassay, Wiley periodical, Inc. (24) 1178-1187