

Advances in Theory of Neural Network and Its Application

Bahman Mashood¹, Greg Millbank²

¹1250 La Playa Street, 304, San Francisco, USA

²At Praxis Technical Group Inc., Nanaimo, Canada

Email: b_mashood@hotmail.com, Gregm@Praxistech.com

Received 12 March 2016; accepted 21 May 2016; published 24 May 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this article we introduce a large class of optimization problems that can be approximated by neural networks. Furthermore for some large category of optimization problems the action of the corresponding neural network will be reduced to linear or quadratic programming, therefore the global optimum could be obtained immediately.

Keywords

Neural Network, Optimization, Hopfield Neural Network, Linear Programming, Cohen and Grossberg Neural Network

1. Introduction

Many problems in the industry involved optimization of certain complicated function of several variables. Furthermore there are usually set of constrains to be satisfied. The complexity of the function and the given constrains make it almost impossible to use deterministic methods to solve the given optimization problem. Most often we have to approximate the solutions. The approximating methods are usually very diverse and particular for each case. Recent advances in theory of neural network are providing us with completely new approach. This approach is more comprehensive and can be applied to wide range of problems at the same time. In the preliminary section we are going to introduce the neural network methods that are based on the works of D. Hopfield, Cohen and Grossberg. One can see these results at (section-4) [1] and (section-14) [2]. We are going to use the generalized version of the above methods to find the optimum points for some given problems. The results in this article are based on our common work with Greg Millbank of praxis group. Many of our products used neural network of some sort. Our experiences show that by choosing appropriate initial data and weights we are able to approximate the stability points very fast and efficiently. In section-2 and section-3, we introduce the

extension of Cohen and Grossberg theorem to larger class of dynamic systems. For the good reference to linear programming, see [3], written by S. Gass. The appearance of new generation of super computers will give neural network much more vital role in the industry, machine intelligent and robotics.

2. On the Structure and Application of Neural Networks

Neural networks are based on associative memory. We give a content to neural network and we get an address or identification back. Most of the classic neural networks have input nodes and output nodes. In other words every neural networks is associated with two integers m and n . Where the inputs are vectors in R^n and outputs are vectors in R^m . neural networks can also consist of deterministic process like linear programming. They can consist of complicated combination of other neural networks. There are two kind of neural networks. Neural networks with learning abilities and neural networks without learning abilities. The simplest neural networks with learning abilities are perceptrons. A given perceptron with input vectors in R^n and output vectors in R^m , is associated with treshhold vector $(\theta_1, \theta_2, \dots, \theta_m)$ and $m \times n$ matrix $(w_{i,j})$. The matrix W is called matrix of synaptical values. It plays an important role as we will see. The relation between output vector

$O = (o_1, o_2, \dots, o_m)$ and input vector vector $S = (s_1, s_2, \dots, s_n)$ is given by $o_i = g\left(\sum_{k=1}^{k=n} w_{i,k} s_k - \theta_i\right)$, with g a

logistic function usually given as $g(x) = \tanh(\beta x)$ with $1 > \beta > 0$ This neural network is trained using enough number of corresponding patterns until synaptical values stabilized. Then the perceptron is able to identify the unknown patterns in term of the patterns that have been used to train the neural network. For more details about this subject see for example (Section-5) [1]. The neural network called back propagation is an extended version of simple perceptron. It has similar structure as simple perceptron. But it has one or more layers of neurons called hidden layers. It has very powerful ability to recognize unknown patterns and has more learning capacities. The only problem with this neural network is that the synaptical values do not always converge. There are more advanced versions of back propagation neural network called recurrent neural network and temporal neural network. They have more diverse architect and can perform time series, games, forecasting and travelling salesman problem. For more information on this topic see (section-6) [1]. Neural networks without learning mechanism are often used for optimizations. The results of D.Hopfield, Cohen and Grossberg, see (section-14) [2] and (section-4) [1], on special category of dynamical systems provide us with neural networks that can solve optimization problems. The input and out put to this neural networks are vectors in R^m for some integer m . The input vector will be chosen randomly. The action of neural network on some vector $X_1 \in R^m$ consist of inductive applications of some function $f: R^m \rightarrow R^m$ which provide us with infinite sequence $X_1, X_2, \dots, X_n, \dots$ where $X_n = f(X_{n-1}) = f^{n-1}(X_1)$. And output (if exist) will be the limit of of the above sequence of vectors. These neural networks are resulted from digitizing the corresponding differential equation and as it is has been proven that the limiting point of the above sequence of vector coincide with the limiting point of the trajectory passing by X_1 . Recent advances in theory of neural networks provide us with robots and comprehensive approach that can be applied to wide range of problems. At this end we can indicate some of the main differences between neural network and conventional algorithm. The back propagation neural networks, given the input will provide us the out put in no time. But the conventional algorithm has to do the same job over and over again. On the other hand in reality the algorithms driving the neural networks are quite massy and are never bug free. This means that the system can crash once given a new data. Hence the conventional methods will usually produce more precise outputs because they repeat the same process on the new data. Another defect of the neural networks is the fact that they are based on gradient descend method, but this method is slow at the time and often converge to the wrong vector. Recently other method called Kalman filter (see (section-15.9) [2]) which is more reliable and faster been suggested to replace the gradient descend method.

3. On the Nature of Dynamic Systems Induced from Energy Functions

In order to solve optimization problems using neural network machinery we first construct a corresponding energy function E , such that the optimum of E will coincide with the optimum point for the optimization problem. Next the energy function E , that is usually positive will induce the dynamic system L_E . The trajectories of L_E will converge hyperbolically to local optimums of our optimization problem. Finally we construct the neural network NN_E which is the digitized version of L_E , where depending on initial points it will converge to some local optimum. As we indicated in section-1, certain category of dynamic system which is

called Hopfield and its generalization which is called Cohen and Grossberg dynamic system will induce a system of neural networks that are able to solve some well known NP problems. More recently the more advanced dynamic systems based on generalization of the above dynamic systems been used in [4] to solve or prove many interesting problems including four color theorem. In the following sequence of lemmas and theorems, we are going to show that if the dynamic system L satisfies certain commuting condition, then it can be induced from an energy function E_L , which is not usually positive and all its trajectories converge hyperbolically to a corresponding attractor points. Furthermore the attractor corresponding to the global optimum is located on the non trivial trajectory. Note that the energy function that is induced from optimization scenario is always positive and the corresponding dynamic system L_G , is a commuting dynamic system.

Suppose we are given dynamic system L , as in the following,

$$dx_i/dt = Q_i(x), i = 1, 2, \dots, n, (x_1, x_2, \dots, x_n) = x \in R^n$$

Also the above equation can be expressed as

$$dx/dt = Q(x).$$

Definition 2.1 We say that the above system L satisfies the commuting condition if for each two indices $i_k \leq n, k = 1, 2, i_1 \neq i_2$ we have

$$\partial Q_{i_1} / \partial x_{i_2} = \partial Q_{i_2} / \partial x_{i_1}.$$

This is very similar to the properties of commuting squares in the V.Jones index theory [5].

The advantage of commuting system as we will show later is that each trajectory $X_{x_0}(t)$ passing through an initial point $x_0 \in R^n$, will converge to the critical point x_∞ , and x_∞ is asymptotically stable.

In particular note that if the dynamic system is induced from an energy function E , then the induced neural network N_E , is robot and stable. In the sense that beginning from one point $x_0 \in R^n$, the neural network will asymptotically will converge to a critical point. This property plus some other techniques make it possible to find the optimum value of E . The following lemma will lead us to the above conclusions.

Lemma 2.2 Suppose the dynamic system L has a commuting property. Then there exists a function $E_L(x)$ acting on R^n such that for every integer $i \leq n$ we have

$$\partial E_L / \partial x_i = -dx_i/dt = -Q_i(x).$$

Furthermore for every trajectory X_{x_0} , $dE_L/dt \leq 0$ and $dE_L/dt = 0$ only on the corresponding critical point x_∞ on which $dE_L/dt = 0 = Q(x_\infty)$.

Proof Let us pick up an integer $i \leq n$. Next define $E_L = -\int Q_i(x) dx_i$. Then for any integer $j \leq n, j \neq i$, we have $\partial E_L / \partial x_j = -\int (\partial Q_i / \partial x_j) dx_i = -\int (\partial Q_j / \partial x_i) dx_i = -Q_j = -dx_j/dt$. finally we have,

$$dE_L/dt = \sum_{i=1}^{i=n} (\partial E_L / \partial x_i) dx_i/dt = -\sum_{i=1}^{i=n} (dx_i/dt)^2 \leq 0.$$

And the equality holds, *i.e.* $dE_L/dt = 0$ only on the critical point x_∞ on which $dE_L/dt = 0$. Q.E.D.

But the problem is that the induced E_L in Lemma 2.2 is not always a positive function. In the case that E_L is a positive function we have the following lemma.

Lemma 2.3 Following the notation as in the above suppose L is a commuting system and that $E_L \geq 0$ is analytical function. Next let $x_\infty \in R^n$ be the critical point for the system L which is on non trivial trajectory with $E_L(x_\infty) = 0$. Then x_∞ is asymptotically stable.

Proof. Following the definition of Liapouov function and using Lemma 2.2, the fact that $E_L(x_\infty) = 0$ implies that regarding to the trajectory passing through x_∞ , it is asymptotically stable. Q.E.D.

There are some cases that we can choose E_L to be a positive function as we will show in the following lemma.

Lemma 2.4 Keeping the same notations as in the above, suppose that there exists a number $\alpha \in R$, such that $\alpha \leq E_L(x)$ for all $x \in R^n$. Then there exists a positive energy function F_L for the dynamic system L .

Proof. Let us define F_L acting on R^n by $F_L(x) = E_L(x) - \alpha$. Then F_L is a positive function. Furthermore $\partial E_L / \partial x_i = -dx_i/dt = Q_i$ and over the trajectories $dE_L/dt \leq 0$, it is equal to zero only over the attracting points. This will complete the proof of the lemma. Q.E.D.

Lemma 2.5 Let L be a commuting dynamic system. Let $E_L \geq 0$ be the induced energy function. Then if $x_\infty \in R^n$ is a point on a non trivial trajectory on which E_L , achieves an optimum α , then the trajectory passing through x_∞ will converge asymptotically to x_∞ .

Proof Set $F_L(x) = E_L(x) - \alpha$, then using Lemma 2.4 E_L is an energy function for L and since $F_L(x_\infty) = 0$ we are done. Q.E.D.

Suppose L is a commuting dynamic system. Let $E_L \geq 0$ be an induced energy function. The main goal of the corresponding neural network N_L is to reach a point $x_\infty \in R^n$ at which E_L will get its optimum value. In Lemma 2.5, we proved that any non trivial trajectory passing through x_∞ will converges to x_∞ asymptotically. In the following we show that in general the above property holds for any attracting point of commuting dynamic systems.

Lemma 2.6 Suppose L is a commuting dynamic system and x_∞ an attractive point. Then the trajectory passing through x_∞ will converge asymptotically to x_∞ .

Proof Let $\delta = E_L(x_\infty)$. Next define the function F_L acting on R^n , by $F_L(x) = (E_L(x) - \delta)^2$. In order to complete the proof we have to show that with regard to the trajectory $X(t, x_\infty)$ passing through x_∞ , $F_L(x)$ is a Liapunov function. For this it is enough to show that $dF_L/dt < 0$. But

$dF_L/dt = d(E_L(x) - \delta)^2/dt = 2(E_L(X(t, x_\infty) - \delta))dE_L/dt$. Finally the fact that $E_L(X(t, x_\infty) - \delta) > 0$ implies that $dF_L/dt < 0$ and this complete the proof. Q.E.D.

Suppose L is a commuting dynamic system and x_∞ is a point on some non trivial trajectory at which E_L reaches its infimum. We want to find a conditions that guarantees the existence of a non trivial trajectory passing through x_∞ .

Definition 2.6 Keeping the same notation as in the above, for a commuting dynamic system L , we call $E_L \geq 0$ canonical if for each critical point x_∞ , with $E_L(x_\infty) = 0$, we have $\lim_{x \rightarrow x_\infty} E_L(x) / \sum_i Q_i^2 < \infty$.

Before proceeding to the next theorem let us set the following notations.

Let $C_\delta = (x \in R^n, \|x\| = \delta)$ and $O_\delta = (x \in R^n, \|x\| \leq \delta)$. Furthermore for $x_\epsilon \in R^n$, let $x_{\epsilon, \delta} \in C_\delta$ be the first point at which the trajectory $X(t, x_\epsilon)$ will intersect C_δ .

Lemma 2.7 Following the above notations suppose without loss of generality, $x_\infty = 0$, and that there is no trajectory passing through a point $x \neq x_\infty$ and converging to x_∞ . Then there exists $\rho > 0$ and a sequences $(\epsilon_i \rightarrow 0), i = 1, 2, \dots$ and $(x_{\epsilon_i} \in O_{\epsilon_i}, i = 1, 2, \dots)$, such that for each $i \in N$, $X(t, x_{\epsilon_i})$ will intersect C_ρ first time at the point $x_{\epsilon_i, \rho} \in C_\rho$.

Proof. Otherwise for each $\rho > 0$, there exists a positive number $\epsilon_\rho > 0$ such that for any positive number $\epsilon < \epsilon_\rho$, and a point $x_\epsilon \in O_\epsilon$, the trajectory $X(t, x_\epsilon)$ will lie in O_ρ . This implies the existence of the sequence $(y_j), j = 1, 2, \dots$ converging to x_∞ , such that for each $j \in N$, $\partial E_L(y_j) / \partial x_i = 0, i \in N$. hence assuming that E_L is analytic this implies that $E_L(x) = 0$ which is a contradiction. Q.E.D.

Theorem 2.8 Keeping the same notation as in the above, suppose we have a commuting system L , with the induced energy function E_L , being canonical. Suppose there exists a point $x_\infty \in R^n$, with $E_L(x_\infty) = 0$. Then there exists a non trivial trajectory $X(t)$, converging to x_∞ .

Proof. Suppose there is no non trivial trajectory passing through x_∞ . Next for any $\epsilon > 0$, let us choose $x_\epsilon \in R^n$, with $\|x_\epsilon - x_\infty\| < \epsilon$. Furthermore let $X(t, x_\epsilon)$, be the trajectory through the point x_ϵ . Now let ρ , and $\epsilon_i, i = 1, 2, \dots$ be as indicated at Lemma 2.7. Next for $j \in N$, consider $X(t, x_{\epsilon_j})$, passing through x_{ϵ_j} . Let

$x_{\epsilon_j, \rho}$ be the first point at which $X(t, x_{\epsilon_j})$ intersect C_ρ . We have,

$dE_L(X(t, x_{\epsilon_j})) / dt = \sum_i Q_i^2(X(t, x_{\epsilon_j})) = G(X(t, x_{\epsilon_j}))$. Now consider the trajectory $X(t, x_{\epsilon_j})$. Let us denote by

$t_{\epsilon_j, \rho}$, the time at which the trajectory $X(t, x_{\epsilon_j})$ arrives at $x_{\epsilon_j, \rho}$. This implies,

$$E_L(X(t_{\epsilon_j, \rho}, x_{\epsilon_j})) - E_L(X(0, x_{\epsilon_j})) = E_L(x_{\epsilon_j, \rho}) - E_L(x_{\epsilon_j}) = \int_{t=0}^{t=t_{\epsilon_j, \rho}} G(X(t, x_{\epsilon_j})) dt.$$

Thus, applying mean value theorem implies

$$E_L(x_{\epsilon_j, \rho}) - E_L(x_{\epsilon_j}) = G(X(\tau, x_{\epsilon_j})) t_{\epsilon_j, \rho}, \text{ with } \tau \in (0, t_{\epsilon_j, \rho}). \quad (1)$$

To complete the proof of the Theorem 2.8, we need the following lemma.

Lemma 2.9 Keeping the same notations as in the above then there exists a $\omega > 0$ such that for $j \in N$ large enough, there exists a positive number $\alpha > 0$, with the property that for every $\epsilon_j \in (\epsilon_i), j$, large enough, $0 < t_{\epsilon_j, \omega} < \alpha$.

Proof. Otherwise there exists a sequence of numbers $(\omega_k), k = 1, 2, \dots$ such that $\omega_k \rightarrow 0$ as $k \rightarrow \infty$, where for each $k \in N$, we have $t_{\epsilon_i, \omega_k} \rightarrow \infty$ as $i \rightarrow \infty$. Thus considering the Equation (1), and the fact That for the indices i, j large enough we have $\|X(t, x_{\epsilon_j})\| < \omega_j, t \in (0, t_{\epsilon_i, \omega_j})$, we can write,

$$t_{\epsilon_i, \omega_j} = E_L(X(t_{\epsilon_i, \omega_j}, x_{\epsilon_i})) / G(X(\tau, x_{\epsilon_i})), \\ \tau \in (0, t_{\epsilon_i, \omega_j}).$$

This using the fact that E_L is canonical will lead to contradiction. Q.E.D

To complete the proof of theorem 2.8, note that for every point x_{ϵ_i} , the points located on the trajectory $X(t, \epsilon_i)$, can be expressed as a continuous function $\Phi(x_{\epsilon}, t)$, of x_{ϵ} and t . The function $\Phi(x_{\epsilon}, t)$ acts on the set $C_{\omega} \otimes [0, \alpha]$, which is a compact set. Hence there exists a number *Length* such that

$L_{\epsilon_i, \omega} = \text{Length}(X(t, x_{\epsilon_i})) \Big|_{t=0}^{t=t_{\epsilon_i, \omega}} < \text{Length}$ for all $i \in N$. Next for each $k \in N$ we divide $X(t, x_{\epsilon_i})$ into 2^k part each of length equal to $L_{\epsilon_i, \omega} / 2^k$. Let us define the following points, $p_{i,k,1} = x_{\epsilon_i}, p_{i,k,2}, \dots, p_{i,k,m}, \dots, p_{i,k,2^k}$ be the points corresponding to the above partitions. Furthermore consider the following countable set of points.

$Q = (p_{i,k,m})_{i,k,m}, i > i_0, k > 1, m = 1, 2, \dots, 2^k$. Furthermore for each triple (i, k, m) , let us define a point $p_{k,m}$ to be a limit point of the set $(p_{i,k,m})_i, i > 1$. Finally consider the following set $S = (p_{k,m})_{k,m}, k > 1, m = 1, 2, \dots, 2^k$. Then the set S^- the closure of S in R^n is an non trivial trajectory passing through x_{∞} . But this is a contradiction to our assumption. Q.E.D.

Lemma 2.10 Keeping the same notations as in the above, suppose for a given commuting system L the induced energy function E_L is positive function. Then E_L is canonical.

Proof. If E_L is not canonical then there exist an increasing sequences of positive numbers, $(\alpha_i) \rightarrow \infty$ as $i \rightarrow \infty$, and $(\epsilon_i) \rightarrow 0$ as $i \rightarrow \infty$, such that for each $i \in N$, there exists a point $u^i \in R^n$, at the distance of less than ϵ_i from u_{∞} with $E_L(u^i) / \sum_{k=1}^{k=n} Q_k^2(u_i) = \alpha_i$.

Next we can assume that there exists a line $U(t)$ in R^n connecting sequence of points u^i to the limiting point u_{∞} , such that $U(1/\epsilon_i) = u^i$. But $E_L = \int Q_i(u) du_i = \int Q_j(u) du_j$, thus using Hopital lemma,

$$\chi = \lim_{t \rightarrow \infty} E_L(u) / \sum_i Q_i^2 = \lim_{t \rightarrow \infty} d(E_L(u)) / dt / d(\sum_i Q_i^2) / dt.$$

Let us set $g(u) = \sum_i Q_i^2$. Then by the above, $\lim_{t \rightarrow \infty} E_L(u) / \sum_i Q_i^2 = \lim_{t \rightarrow \infty} g(u) / g'(u)$. Next if $\lim_{t \rightarrow \infty} g(u) / g'(u) = 0/0$, then using Hopital lemma we have, $\chi = \lim_{t \rightarrow \infty} g'(u) / (g(u))''$. continue this process suppose using induction that $\chi = \lim_{t \rightarrow \infty} (d^m g(u) / dt^m) / (d^{m+1} g(u) / dt^{m+1}) = 0/0$, then we get by Hopital lemma, $\chi = \lim_{t \rightarrow \infty} (d^{m+1} g(u) / dt^{m+1}) / (d^{m+2} g(u) / dt^{m+2})$.

Now if for each $m \in N$, $\lim_{t \rightarrow \infty} (d^m g(u) / dt^m) = 0$, then using the fact that $g(u(t))$ is analytic function we get that $g(t) = 0$. which is a contradiction. Hence there exists an integer $m \in N$, such that

$\lim_{t \rightarrow \infty} (d^m g(u) / dt^m) = 0$ but $\lim_{t \rightarrow \infty} (d^{m+1} g(u) / dt^{m+1}) \neq 0$. Therefore the above arguments imply that

$\chi = \lim_{t \rightarrow \infty} E_L(u) / \sum_i Q_i^2 = 0$, and this is a contradiction to the assumption. Q.E.D.

As a result of the Lemma 2.10 we get that if $E_L \geq 0$ the system L is always canonical, hence we have the the following corollary, Corollary 2.11 Keeping the same notations as in the above, The results of Theorem 2.8 holds as long as $E_L \geq 0$.

At this point we have to mention that non trivial trajectories will supply us with much more chance of hitting the global optimum, once we perform a random search to locate it.

For some dynamic systems L which is expressed in the usual form $dx/dt = Q(x)$, the commuting condition does not hold. For example consider the Hopfield neural network and its corresponding dynamic system,

$$du_i/dt = u_i + \sum_{j=1}^{j=n} w_{i,j} g(u_j), i = 1, 2, \dots, n,$$

where $g(x) = \tanh(x)$. It is clear that the above system does not posses commuting properties. Let us multiply both side of the i'th equation in the above by $dg(u_i)/du_i = g'(u_i)$, to get the dynamic system L_1 , given in the following as, $(du_i/dt) g'(u_i) = g'(u_i) \left(-u_i + \sum_{j=1}^{j=n} w_{i,j} g(u_j) \right)$, therefore we get,

$$dg(u_i)/dt = (1 - g^2(u_i)) \left(-u_i + \sum_{j=1}^{j=n} w_{i,j} g(u_j) \right) \tag{2}$$

Let us denote $u_{1,i} = g(u_i)$ and $Q_{1,i} = du_{1,i}/dt$, the the dynamic system L_1 can be expressed as

$$du_{1,i}/dt = Q_{1,i}.$$

It is clear that if $U_i(t)$ is a trajectory of the system L , and provided $g'(u_i) \neq 0$, then $U_i(t)$ is a trajectory of system L_1 too. Now the commuting property holds for the right side of Equation (2), hence using the same techniques as before we can construct the energy function $E(u)$ for the system (2) such that, $\partial E(u)/\partial u_i = -Q_{1,i}$, furthermore we have

$$dE(u(t))/dt = -\sum_{i=1}^{i=n} (1 - g^2(u_i)) (du_i/dt)^2.$$

This implies that $dE/dt < 0$ except on the attractive points. Now by the results of the Lemma 2.3, this implies that for any trajectory $U(t, u_0)$, the convergent asymptotically to the corresponding attractive point u_∞ .

As we mentioned before more generalized version of Hopfield dynamic system which is called Cohen and Grossberg dynamic system is given as in the following.

$$du_i/dt = a_i(u_i) \left(b_i(u_i) - \sum_{j=1}^{j=n} c_{i,j} g_j(u_j) \right), i = 1, 2, \dots, n \tag{3}$$

where the set of coefficients $c_{i,j}, i, j \in (1, 2, \dots, n)$, will satisfy, $c_{i,j} = c_{j,i}$. Furthermore $a_i(u_i) \geq 0, g_i(u_i) \geq 0, i = 1, 2, \dots, n$.

Likewise the system (2), system (3) is not a commuting system. But if we multiply both side of the ith equation in system (3) by $g'_i(u_i)$ then we get a dynamic system where its right side is commuting. Hence each of the trajectories converge asymptotically to corresponding attracting point.

4. Reduction of Certain Optimization Problems to Linear or Quadratic Programming

In solving optimization problems using neural network we first form an energy function $E(u)$, corresponding to the optimization problem. Next the above energy will induces the dynamic system L that its trajectories converge to local optimum solutions for the optimization problem.

Given the energy function $E(u)$ acting on R^n , the induced dynamic system L is given in the following,

$$du_i/dt = -\partial E/\partial u_i = Q_i.$$

As we showed L is a commuting system.

As an example consider the travelling salesman problem. As it has been expressed in section 4.2, page 77 of [1] the energy function E is expressed as in the following,

$$E = 0.5 \left(\sum_{i,j,k} d_{i,j} u_{i,k} (u_{j,k+1} + u_{j,k-1}) \right) + 0.5\gamma \left(\sum_{k=1}^{k=n} \left(1 - \sum_{i=1}^{i=n} u_{i,k} \right)^2 + \sum_{i=1}^{i=n} \left(1 - \sum_{k=1}^{k=n} u_{i,k} \right)^2 \right)$$

where we are represent the points to be visited by travelling salesman as $(p_i), i = 1, 2, \dots, n$ and the distance of

point p_i to the point p_j by $d_{i,j}$. Since in solving optimization problems using neural network we are interested in variables that are of 0 or 1 nature therefore let us define new set of variables by,

$$(v_{j,k}), j, k = 1, 2, \dots, n, \quad v_{j,k} = 1/2(\tanh(\lambda_0 u_{j,k}) + 1).$$

Thus the dynamic system L corresponding to E , can be written as in the following,

$$\partial E / \partial u_{j,k} = -dv_{j,k} / dt = -\left(\sum_{l=1}^{l=n} d_{j,l} \gamma (u_{j,l} / d_{j,l} + u_{l,k+1} / \gamma + u_{l,k} / d_{j,l} + u_{l,k-1} / \gamma) - 2\gamma\right)$$

As we proved the point at which E reaches its optimum is a limiting point u_∞ , of a trajectory belonging to the dynamic system L . as given in the above.

Next as we had,

$$dv_{j,k} / dt = -\partial E / \partial u_{j,k} = \lambda_0 / 2 (1 - \tanh^2(\lambda_0 u_{j,k})) dy_{j,k} / dt = \sum_{l=1}^{l=n} d_{j,l} (y_{j,l} / d_{j,l} + y_{l,j+1} + y_{l,j} / d_{j,l} + y_{l,j-1} / \gamma).$$

But $du_{j,k} / dt = (\lambda_0 / 2) (1 - \tanh(\lambda_0 v_{j,k})) dv_{j,k} / dt$ hence we get the following dynamic system L_1 ,

$$du_{j,k} / dt = \left(\sum_{l=1}^{l=n} d_{j,l} \gamma (u_{j,l} / d_{j,l} + u_{l,k+1} / \gamma + u_{l,k} / d_{j,l} + u_{l,k-1} / \gamma) - 2\gamma\right).$$

Thus as we showed the system L_1 is a commuting system and its critical points coincides with the critical points of the system L .

Now consider the following set of variables $y_{i,j}, i, j = 1, 2, \dots, n$, $y_{i,j} = u_{i,j} - q_{i,j}$, where we choose the set $q_{i,j}, i, j = 1, 2, \dots, n$. to satisfy,

$$\sum_{l=1}^{l=n} d_{j,l} (q_{j,l} / d_{j,l} + q_{l,s+1} / \gamma + q_{l,s} / d_{j,l} + q_{l,s-1} / \gamma) = 2, j, s = 1, 2, \dots, n.$$

Hence we will have the following set of equations,

$$\sum_{l=1}^{l=n} (q_{j,l} + d_{j,l} q_{l,s+1} / \gamma + q_{l,s} + q_{l,s-1} d_{j,l} / \gamma) = 1, j, s = 1, 2, \dots, n.$$

Replacing $u_{j,k}$ by $y_{j,k}$ in equations of system L_1 implies,

$$dy_{j,k} / dt = -\left(\sum_{l=1}^{l=n} d_{j,l} (y_{j,l} / d_{j,l} + y_{l,s+1} + y_{l,s} / d_{j,l} + y_{l,s-1} / \gamma)\right).$$

But $y_{j,k} \rightarrow 0$ as, $t \rightarrow \infty$, which implies that, $u_{j,k} \rightarrow q_{j,k}$ as $t \rightarrow \infty$, hence this implies that the above equations together with the optimality of the expression $F = \sum_{i,j,k=1}^{i,j,k=n} d_{i,j} u_{i,k} (u_{j,k+1} + u_{j,k-1})$ is a system of quadratic linear programming that will give us the optimum value much faster than usual neural network. In fact this method can be applied to many types of optimization problems which guarantee fast convergence to desired critical point.

Let us consider the Four color Theorem. The similar scenario to Four color Theorem is to consider the have two perpendicular axis X and Y and sets of points $P_1 = (x_i), i = 1, 2, \dots, n$ on X axis and the set of points $P_2 = (y_j), j = 1, 2, \dots, n$ on the Y axis. with $x_i = (i/n, 0)$ and $y_j = (0, j/n)$. Let us set the following four points $O = (0, 0), A = (0, 1), B = (1, 0), C = (1, 1)$. We want to connect the points of the sets P_1 to P_2 such that the connecting line will intersect only outside the square $OACB$. Now suppose we take a point $p_1 = (i_1/n, 0)$ and $q_1 = (0, j_1/n)$, $p_2 = (i_2/n, 0)$ and $q_2 = (0, j_2/n)$. Next connecting the points p_1 and q_1 will give us the line Y_1 . Connecting the points p_2 and q_2 will give us the line Y_2 . We want to choose the lines Y_1 and Y_2 such that the point z which is the intersection of Y_1 and Y_2 will stay outside the square $OABC$. Suppose $z = (z_1, z_2)$. Then the above condition is equivalent to the fact that either $z_1 \geq 1$ or $z_1 \leq 0$. But

$z_1 = (j_2 - j_1) / n (j_1 / i_1 - j_2 / i_2)$. Now the above condition is equivalent to the following set of inequalities,

Case-1. If $j_2 \geq j_1$, then either $j_2 - j_1 \geq n (j_1 / i_1 - j_2 / i_2)$ or $j_2 \geq i_2 j_1 / i_1$. or symmetrically,

Case-2. If $j_1 \geq j_2$, then either $j_1 - j_2 \geq n (j_2 / i_2 - j_1 / i_1)$ or $j_1 \geq i_1 j_2 / i_2$.

Next let us take the set of following variables, $(w_{i,0,k})$ and $(w_{0,j,k})$, with $i, j \in (1, 2, \dots, n)$ and $k \in (1, 2, \dots, m)$. Furthermore consider the following equalities,

$$\left(\sum_{i,0,k} w_{i,0,k}\right)_{k=1}^{k=m} = 1, \text{ for all } i \in (1, 2, \dots, n).$$

$$\left(\sum_{0,j,k} w_{0,j,k}\right)_{k=1}^{k=m} = 1, \text{ for all } j \in (1, 2, \dots, n).$$

Furthermore let us set the following energy function to be optimize,

$$E = \sum_{i,j,k_1} d_{i,j} w_{i,0,k_1} + \sum_{k_2 \neq k_1} w_{0,j,k_2} + \lambda \left(\sum_{i=1}^{i=n} \left(1 - \sum_{k=1}^{k=m} w_{i,0,k}\right)^2 \right) + \left(\sum_{j=1}^{j=n} \left(1 - \sum_{k=1}^{k=m} w_{0,j,k}\right)^2 \right)$$

also we must meet the conditions of Case-1 and Case-2 over the indices.

Now the above system is equivalent to find the optimum solution for coloring.

So as before assuming, $W = 1/2(\tanh(\lambda_0 V) + 1)$ we have, $\partial E / \partial w_{i,0,k} = -d/dt(v_{i,0,k})$ and $\partial E / \partial w_{0,j,k} = -d/dt(v_{0,j,k})$.

This implies, $d/dt(v_{i,0,k}) = \left(-\sum_{k=1}^{k=m} w_{i,0,k} - c_i / 2\lambda\right) 2\lambda$ with $c_i = \sum_{j=1}^{j=m} d_{i,j}$.

Next let us define $w_{i,0,k} = y_{i,k} + q_{i,k}$

On the other hand it is easy to show that $d/dt(v_{0,j,k}) = 2\lambda \left(-\sum_{k=1}^{k=m} w_{0,j,k} - d_j / 2\lambda\right)$ with $d_j = \sum_{i=1}^{i=n} d_{i,j}$.

Next let us define, $w_{0,j,k} = z_{j,k} + p_{j,k}$.

Finally if the following inequalities holds,

$$\sum_{k=1}^{k=m} q_{i,k} = -c_i / 2\lambda, i = 1, 2, \dots, n$$

$$\sum_{k=1}^{k=m} p_{j,k} = -d_j / 2\lambda, j = 1, 2, \dots, n$$

where the indices satisfy, the conditions of Case-1 and Case-2. then as t tends to infinity we have, $w_{i,0,k} = q_{i,k}$ and $w_{0,j,k} = p_{j,k}$.

At this point using the above arguments it is enough to find Q , and P , satisfying the above equalities and will optimize the following expression,

$$\sum_{i,j,k_1} d_{i,j} \left(q_{i,k_1} + \sum_{k_2 \neq k_1} p_{j,k_2} \right), i, j = 1, \dots, n \text{ and } k_1 = 1, 2, \dots, m.$$

Therefore the above equations together with optimization expression will form a system of linear programming that will converge to the optimal solution at no time.

5. Conclusion

In this article we introduced the methods of approximating the solution to optimization problems using neural networks machinery. In particular we proved that for certain large category of optimization problems the application of neural network methods guaranties that the above problems will be reduced to linear or quadratic programming. This will give us very important conclusion because the solution of the optimization problems in these categories can be reached immediately.

References

- [1] Hetz, J., Krough, A. and Palmer, R. (1991) Introduction to the Theory of Neural Computation. Addison Wesley Company, Boston.
- [2] Haykin, S. (1999) Neural Networks: A Comprehensive Foundation. 2nd Edition, Prentice Hall, Inc., Upper Saddle River.
- [3] Gass, S. (1958) Linear Programming. McGraw Hill, New York.
- [4] Yoshiyasu, T. (1992) Neural and Parallel Processing. The Kluwer International Series in Engineering and Computer Science: SEeS0164.
- [5] Jones, V. and Sunder, V.S. (1997) Introduction to Subfactors. Cambridge University Press, Cambridge.