

YAPAY SİNİR AĞ TOPOLOJİSİNİN OPTİMUM SAKLI KATMAN AKTİVASYON SEVİYESİ ÜZERİNDEKİ BELİRLEYİCİLİĞİ

Halis ALTUN

Tankut YALÇINOZ

B. Sami TEZEKİCİ

Elektrik-Elektronik Mühendisliği Bölümü

Mühendislik Fakültesi

Niğde Üniversitesi, 51100, NİĞDE

e-posta: halis_altun@ieee.org

e-posta: t.yalcinoz@ieee.org

Anahtar Kelimeler: Yapay Sinir Ağları, Geri Yayılım Algoritması, Optimum Nöral Aktivasyon Seviyesi

ABSTRACT

Eğitim setinin nöral öğrenme üzerindeki etkisi literatürde yapılan çalışmalar ile gösterilmiştir. Literatürde bu yönde yapılan çalışmalar sonucunda eğitim setinin optimum sunumu için değişik uygulamalar önerilmiştir. Bu çalışmada, YSA topolojisinin, ikili (binary) eğitim veri setinin, saklı katman nöronları için optimum aktivasyon seviyesini nasıl belirlediği ve bu doğrultuda daha hızlı bir nöral eğitim için ikili eğitim veri setinin ortalama değerinin nasıl modifi edilmesi gerektiği araştırılmıştır. Elde edilen sonuçlar saklı katman optimum aktivasyon seviyesinin giriş ve çıkış katmanlarındaki nöron sayısına bağımlılığını ortaya koymuştur. Bu sonuç saklı katman nöron sayısının nöral öğrenme süreci üzerinde belirleyici etkisinin kısıtlı olduğunu gösteren literatürdeki genel yaklaşımlar ile de uyumludur.

1. GİRİŞ

Yapay Sinir Ağlarında (YSA) nöral öğrenimin optimize edilmesi problemi literatürde oldukça yaygın olarak incelenmiştir. Özellikle yaygın bir kullanım alanı bulan Multi-Layered Perceptron (MLP) Tipi Çok Katmanlı Feed-forward YSA'larının optimizasyonu noktasında oldukça geniş bir literatür oluşturulmuştur. Bu çalışmalarda öğrenme algoritması, eğitim seti veya ağ topolojisinin optimize edilmesi amaçlanmıştır [1-5]. Eğitim setinin öğrenme üzerindeki etkisi, Yapay Sinir Ağları üzerinde çalışan araştırmacılar tarafından uzun bir zamanı bilinmektedir. Bu doğrultuda giriş eğitim verilerinin kodlanması örnek olarak verilebilir. YSA için bir benchmark olarak seçilmiş olan "two-spiral" probleminin giriş eğitim veri setinin ikili kodlanması ile kolayca çözümlenebildiği gösterilmiştir [5-6].

Bu çalışmada YSA topolojisinin saklı katman nöronları için optimum aktivasyon seviyesini nasıl belirlediği araştırılmıştır. Ayrıca ikili (binary) giriş eğitim veri setinin, optimum olarak tespit edilen nöral aktivasyon değerleri göz önüne alınarak nasıl düzenlenmesi gerektiği belirlenmiştir. Elde edilen sonuçlar optimum aktivasyon seviyesinin giriş ve çıkış katmanlarındaki nöron sayısına bağımlılığını göstermektedir. Bu sonuç saklı katman nöron sayısının nöral öğrenme sürecinde belirleyici etkisinin kısıtlı olduğunu gösteren literatürdeki genel

yaklaşımlar ile de uyumludur. Literatürde saklı katman nöron sayısının, optimum bir öğrenme için nasıl belirlenebileceğine ilişkin genel kabul görmüş bir kuralın olmaması da bu çalışmada ortaya konan sonuçlar göz önüne alınarak açıklanabilir.

2. OPTİMUM NÖRAL AKTİVASYON SEVİYELERİ

Bir MLP tipi YSA 'ın eğitimi için Geri Yayılım (GY) algoritması kullanılması durumunda, her bir katman göz önüne alınarak, hangi nöral aktivasyon seviyeleri için algoritma tarafından üretilecek ağırlık yenileme işaretlerinin büyük genliklere ulaşabileceği araştırılabilir. İkili eğitim veri seti için yapılacak bu inceleme literatürde analog veri seti kullanılarak yapılan incelemelerde elde edilen sonuçlar dikkate alınarak analitik ve istatistiksel olarak incelenebilir [7].

Bu inceleme için bir MLP YSA'da giriş, saklı ve çıkış katmanlarında mevcut nöronlar için aktivasyon vektörlerini x_1 , x_b ve x_o olarak tanımlayalım. Giriş katman aktivasyon vektörünün K boyutlu, saklı katman aktivasyon vektörünün L boyutlu ve çıkış katman aktivasyon vektörünün ise M boyutlu olduğu varsayalım. Bu durumda her bir katman aktivasyon vektörü nöron aktivasyon seviyeleri göz önüne alınarak aşağıdaki şekilde verilebilir.

$$x_1^{(s)} = [x_1^{(s)}, x_2^{(s)}, \dots, x_K^{(s)}]^T \quad (1)$$

$$x_b^{(s)} = [x_1^{(s)}, x_2^{(s)}, \dots, x_L^{(s)}]^T \quad (2)$$

$$x_o^{(s)} = [x_1^{(s)}, x_2^{(s)}, \dots, x_M^{(s)}]^T \quad (3)$$

s eğitim setinde mevcut toplam veri örüntü miktarı ve x bir nöron için aktivasyon değerini temsil eder.

MLP YSA'ların eğitiminde oldukça yaygın olarak kullanılan GY algoritması, giriş katman-saklı katman ve saklı katman-çıkış katmanları arasında mevcut olan ağırlıklı bağlantıların, w_{1b} ve w_{bo} , optimize edilerek, hedef örüntü vektörü t_o ve çıkış katman aktivasyon vektörü x_o arasındaki hatanın önceden tespit edilmiş belirli bir hatadan daha büyük olmamasını temin edebilmelidir. Bu optimizasyon işlemi başarı ile sonuçlanır ise YSA'nın incelenen problemi öğrenmiş ve bu problem için bir lokal çözüm sunmuş olduğu

kabul edilebilir. Bu durumda aşağıdaki eşitlik yazılabilir.

$$x_o^{(s)} = t_o^{(s)} \pm e \quad s=0,1,\dots,n \quad (4)$$

Burada e önceden belirlenmiş hata tolerans miktarı ve n eğitim setinde mevcut veri miktarı olarak verilmiştir.

Her bir ağırlıklı bağlantı için algoritmanın üreteceği ağırlık yenileme işareti tek tek incelenir ise bağlantıların her bir katman nöral aktivasyon seviyeleri dikkate alınarak verilmesi daha uygun olacaktır. Bu durumda algoritma tarafından üretilen hata yenileme işaret ifadeleri klasik formundan farklı olarak aşağıda olduğu gibi verilebilir.

$$\Delta w_{ho} = -\eta x_o' \Delta x_o x_h \quad (5)$$

$$\Delta w_{ih} = -\eta x_h' \sum_o x_o' w_{ho} \Delta x_o x_i \quad (6)$$

$$\delta_o = x_o (1 - x_o) (t_o - x_o) \quad (7)$$

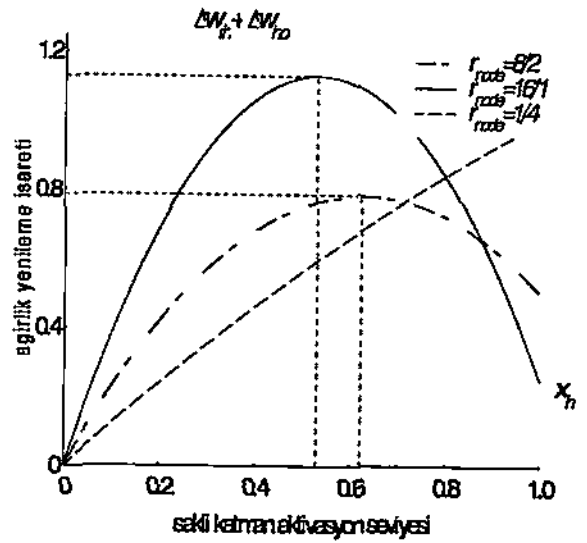
$$\delta_h = x_h (1 - x_h) \sum_o \delta_o w_{ho} \quad (8)$$

$$x_o = f_{sig} \left(\sum_h x_h w_{ho} \right) \quad (9)$$

$$x_h = f_{sig} \left(\sum_i x_i w_{ih} \right) \quad (10)$$

Burada $\Delta x_o = (t_o - x_o)$

- $f_{sig}()$: sigmoid aktivasyon fonksiyonu,
- δ : delta hata ifadesi,
- η : öğrenme oranı,
- i, h, o : giriş, saklı ve çıkış katman indisleri,
- K, L, M : giriş, saklı ve çıkış katman nöron sayıları,
- t_o : hedef değer,
- x_o : çıkış aktivasyon seviyesi,
- x_o' : çıkış aktivasyon seviyesinin türevi,
- x_h : saklı katman aktivasyon seviyesi,
- x_h' : saklı katman aktivasyon seviyesinin türevi,
- x_i : giriş aktivasyon seviyesi,
- w_{ho} : saklı katman-çıkış katmanı arasında ağırlıklı bağlantı,
- Δw_{ho} : saklı katman-çıkış katmanı arasında ağırlıklı bağlantılar için ağırlık yenileme işareti,
- w_{ih} : giriş katmanı-saklı katman arasında ağırlıklı bağlantılar,
- Δw_{ih} : giriş katmanı-saklı katman arasında ağırlıklı bağlantılar için ağırlık yenileme işareti.



Şekil 1. Toplam ağırlık yenileme işaretinin x_h saklı katman aktivasyon seviyesine bağımlı olarak, farklı ağ topolojilerinde, değişimi ve her bir topoloji için optimum x_h değerleri.

Analitik inceleme yapabilmek için 1-1-1 yapısına sahip bir MLP YSA göz önüne alınsın ve ağırlık yenileme işaretlerinin maksimum genlik seviyesine ulaşabilmesi için giriş, saklı ve çıkış katman nöron aktivasyon seviyelerinin ne olması gerektiği araştırılsın. Eğitim veri setinin $[0,1]$ aralığında analog bir veri olduğunu kabul edelim. Bu durumda giriş katman nöron aktivasyon seviyesi, x_i , (5) denkleminde göre aktivasyon düzleminin üst sınırı olan 1 değerine sahip iken algoritma yüksek genlikli bir ağırlık yenileme işaretini giriş katmanı-saklı katman arasında ağırlıklı bağlantılar, Δw_{ih} için üretecektir. Gene (5) ve (6) denklemlerinden, çıkış katman nöron aktivasyon seviyesinin, x_o , bu ifadenin türevinin maksimum seviyeye sahip olduğu aktivasyon düzleminin ortasında seçilmesinin gerekliliği görülmektedir. Ancak saklı katman aktivasyon seviyesi x_h için yukarıdaki denklemlerden hareket ile bir optimum nokta belirlemek mümkün değildir. x_h değerinin aktivasyon düzleminin üst sınırında seçilmesi durumunda (5) denkleminde göre saklı katman-çıkış katmanı arasında ağırlıklı bağlantılar için ağırlık yenileme işaretinin, Δw_{ho} maksimum noktasına ulaşmasına rağmen, aynı değer için (6) nolu denkleme göre giriş katmanı-saklı katman arasında ağırlıklı bağlantılar, w_{ih} için ağırlık yenileme işaretinin üretilmediği görülmektedir. Bu durumda algoritma tarafından üretilen toplam ağırlık yenileme işareti dikkate alınarak x_h için optimum aktivasyon noktası tespit edilmelidir. Bu optimum nokta ağ topolojisinden hareketle ağda mevcut ağırlıklı bağlantılar dikkate alınarak hesaplanabilir. Her hangi bir MLP tipi YSA topolojisinde, $K \cdot L$ sayısınca giriş katmanı-saklı katman arasında, w_{ih} ve $L \cdot M$ sayısınca saklı katman-çıkış katmanı arasında, w_{ho} ağırlıklı bağlantılar mevcuttur. Yukarıda belirtilen durum dikkate alınarak GY algoritması tarafından üretilen

toplam ağırlık yenileme işareti ($K \cdot L / K \cdot M$) oranına bağlı olarak değişen bir büyüklüğe sahip olacaktır. Bu iki ağırlıklı bağlantı sayıları bir orantı olarak, r_{node} verilir ise GY algoritması tarafından üretilen ağırlık yenileme işaretinin x_h saklı katman aktivasyon seviyesine bağımlı olarak nasıl değişeceği Şekil 1'de gösterilmiştir. Burada r_{node} orantısı

$$r_{node} = \frac{K}{M} \quad (11)$$

şeklinde verilmiştir.

GY algoritmasının ağırlık yenileme işaret üretimini maksimum yapacak x_h aktivasyon değerleri, (5) ve (6) denklemlerinden hareket ile giriş ve çıkış katman nöron sayılarına bağımlı olarak

$$x_{h,opt} = \frac{K + M}{2K} \quad (12)$$

şeklinde elde edilir. Bu ifadeden hareket ile elde edilen optimum saklı katman aktivasyon değerleri Tablo 1'de sıralanmıştır.

Tablo 1. K ve M değerlerine bağlı olarak optimum x_h değerleri

K	M	x_h optimum aktivasyon seviyesi
3	1	0,67
3	2	0,83
3	3	1,00
4	1	0,63
4	2	0,75
4	3	0,88
5	1	0,60
5	2	0,70
5	3	0,80
6	1	0,58
6	2	0,67
6	3	0,75
7	1	0,57
7	2	0,64
7	3	0,71
7	4	0,86

3. VERİ KODLAMA

Katmanlarda bulunan nöronların aktivasyon seviyelerinin, eğitim setini oluşturan x_i ve t_o giriş ve hedef vektörleri tarafından oluşturulan ikili veri setinin üzerinde yapılacak kodlama sayesinde nasıl optimum noktaya taşınabileceği incelenir.

x_h saklı katman nöron aktivasyon seviyesi için tespit edilen ve giriş katman nöron sayısı K ve çıkış katman nöron sayısı M değerlerine bağlı olarak ifade edilen optimum nokta, nöral öğrenme sürecini

hızlandıracaktır. Ancak x_i dışında x_h ve x_o aktivasyon değerlerinin GY algoritması tarafından belirleniyor olmasından dolayı nöral aktivasyon seviyelerinin optimum seviyelerine nasıl çekilebileceği problemi ortaya çıkmaktadır.

Literatürde analog bir eğitim veri seti için giriş veri vektörlerinin dağılım karakteristiğinin lineer olmayan ölçeklendirme kullanılarak optimum nöral aktivasyon seviyesine yakın bir ortalama değer alacak şekilde düzenlenebileceği gösterilmiştir [7]. İkili eğitim veri seti için ise giriş vektörü uygun bir kodlama sonucunda her bir patternde '0' görünme miktarı, P(0), ve '1' görünme miktarı P(1), düzenlenerek, giriş veri vektörlerinin dağılım karakteristiği optimum değere yakın bir ortalama değer alacak şekilde ayarlanabilir.

Çıkış aktivasyon seviyelerinin optimum noktaya çekilmesi, bu değerlerin algoritma tarafından belirleniyor olmasından dolayı, doğrudan bir yöntem izlenerek yapılamaz. Ancak algoritmanın çıkıştaki hatayı düşürecek bir şekilde, çıkış aktivasyon seviyelerini hedef vektör, t_o , değerlerine yakınlaştırması dikkate alınır ise t_o hedef veri vektör değerlerinin x_o için optimum olarak belirlenen değerleri alacak şekilde düzenlenmesi ile çıkış aktivasyon vektör dağılım karakteristiği optimum değere yakın bir ortalama değere çekilebilir.

Eğitim seti üzerinde manipülasyon yaparak x_h saklı katman aktivasyon seviyelerinin optimum noktaya taşınması ise yukarıda verilen analitik incelemelere göre mümkün değildir. Ancak analog eğitim veri seti kullanılarak yapılan incelemeler, öğrenme sürecinde saklı katman aktivasyon seviyelerinin, giriş veri vektör dağılımına benzer bir dağılıma sahip olduğunu göstermiştir [7]. Bu durumda giriş veri vektörlerinin x_h için optimum olarak verilen dağılım değerlerine sahip olacak şekilde düzenlenmesinin saklı katman aktivasyon seviyelerinin optimum noktalara yaklaşmasını sağlayacağı beklenebilir.

4. DENEYSEL SONUÇLAR

Yukarıda yapılan incelemeler doğrultusunda ikili eğitim veri seti için elde edilen teorik sonuçların doğrulanması için bir MLP YSA'yı kullanılarak tek ve çift parity (ExOr ve ExNor) problemleri ile deneysel çalışmalar yapıldı. İkili veri setleri değişik dağılım karakteristiklerine sahip olacak şekilde düzenlenerek her bir problem için 3 ayrı eğitim seti oluşturuldu. Bu eğitim setleri için, giriş ve çıkış veri setlerinde '0' ve '1' görünme miktarları, sırasıyla $P_i(0)/P_i(1)$ ve $P_o(0)/P_o(1)$ ve r_{node} değerleri hesaplanarak Tablo II'de verilmiştir.

Aynı şekilde ağ topolojisinin optimum saklı katman aktivasyon değerleri x_h üzerindeki etkisinin doğrulanması için K/M oranını değiştirecek şekilde bir kodlama uygulanmıştır. Bunun için temperature kodlama kullanılarak giriş katman nöron sayısı değiştirilebilir. Temperature kodlama; ikili sayıların desimal karşılığının belirlenmesi ve ikili sayının bu desimal büyüklük miktarı kadar 1'ler ile ifade

edilmesi temeline dayanır. Örneğin 000101 ikili sayısının desimal karşılığı 5'tir ve bu ikili sayı temperature kodlama kullanılarak 011111 şeklinde ifade edilir. Tablo II'de temperature kodlama kullanılarak elde edilen eğitim veri setleri için $P_i(0)/P_i(1)$, $P_o(0)/P_o(1)$ ve r_{mode} değerleri gösterilmiştir. Tablo II aynı zamanda her bir eğitim setinde giriş ve çıkış veri setlerinin ortalama değerlerini de vermektedir.

Tablo II. Oluşturulan eğitim setlerinde hesaplanan $P_i(0)/P_i(1)$ ve $P_o(0)/P_o(1)$ ve r_{mode} değerleri. (her bir eğitim setine ait ortalama değer parantez içerisinde verilmiştir.)

Veri Seti	$P_i(0)/P_i(1)$	$P_o(0)/P_o(1)$	r_{mode}
Kodlamasız			
EXOR1.TXT	24/24 (0.500)	8/8 (0.500)	3/1
EXOR2.TXT	18/30 (0.625)	10/6 (0.375)	3/1
EXOR3.TXT	30/18 (0.375)	6/10 (0.625)	3/1
EXNOR1.TXT	24/24 (0.500)	8/8 (0.500)	3/1
EXNOR2.TXT	18/30 (0.625)	10/6 (0.375)	3/1
EXNOR3.TXT	30/18 (0.375)	6/10 (0.625)	3/1
Temperature Kodlama			
EXOR1T.TXT	56/56 (0.500)	8/8 (0.500)	7/1
EXOR2T.TXT	42/70 (0.625)	10/6 (0.375)	7/1
EXOR3T.TXT	70/42 (0.375)	6/10 (0.625)	7/1
EXNOR1T.TXT	56/56 (0.500)	8/8 (0.500)	7/1
EXNOR2T.TXT	42/70 (0.375)	10/6 (0.375)	7/1
EXNOR3T.TXT	70/42 (0.625)	6/10 (0.625)	7/1

Yapılan kodlamalar sonucu 12 farklı eğitim veri seti oluşturulmuş ve bu setler kullanılarak bir YSA eğitilmiştir. YSA topolojisinde saklı katman nöron sayısı 5 olarak sabit tutulmuş ve kullanılan eğitim setine göre giriş ve çıkış katman nöron sayıları 3/1 ve 7/1 olacak şekilde alınmıştır. Öğrenme oranı (learning rate) 0.3 momentum terimi ise 0.4 seçilmiştir. Her deneme için 30000 olarak belirlenen bir iterasyon sınırı tespit edilmiştir. Her bir eğitim seti için 10 farklı başlangıç noktası seçilerek toplam 120 deneme gerçekleştirilmiştir. Bu denemeler sonucunda her bir eğitim seti için nöral öğrenmenin tamamlanması için gerekli olan ortalama iterasyon sayısı hesaplanmış ve sonuçlar Tablo III'de verilmiştir.

Tablo III. Her bir Eğitim veri seti için YSA'nın öğrenme sürecinde ortalama iterasyon miktarı

Eğitim Veri Seti	Ortalama İterasyon
EXOR1.TXT	1371
EXOR2.TXT	1150
EXOR3.TXT	1839
EXNOR1.TXT	1845
EXNOR2.TXT	1253
EXNOR3.TXT	2328
EXOR1T.TXT	358
EXOR2T.TXT	389
EXOR3T.TXT	382
EXNOR1T.TXT	352
EXNOR2T.TXT	376
EXNOR3T.TXT	381

Sonuçlar eğitim veri seti üzerinde ağ topolojisi tarafından belirlenen r_{mode} değerlerinin ve giriş veri setlerinde '0' ve '1' görünme miktarlarının $P_i(0)/P_i(1)$ nöral öğrenme üzerindeki teorik çalışmaların öngördüğü etkilerini ortaya çıkarmıştır. İkili eğitim setlerinde ortalama değerin, $r_{mode}=3/1$ için optimum saklı katman aktivasyon değeri olarak tespit edilen 0.67'e yakın olmasına bağımlı olarak nöral öğrenmede bir iyileşme görülmektedir. EXOR2 ve EXNOR2 olarak işaretlenen eğitim setleri EXOR1, EXOR3 ve EXNOR1, EXNOR3 olarak işaretlenen diğer eğitim setlerine oranla oldukça hızlı bir nöral öğrenme göstermektedir. Bu ise teorik çalışmanın öngördüğü gibi ağırlıklı bağlantı yenileme işaretlerinin ağ topolojisi ve giriş veri dağılımı karakteristiğine bağımlı olarak değişim gösterdiğini ortaya çıkartmaktadır. Tablo III'den aynı zamanda temperature kodlama kullanılarak yapılan denemelerde $r_{mode}=7/1$ değeri için optimum aktivasyon saklı katman değeri olarak verilen 0.57'e yakın ortalama değere sahip eğitim veri setlerinin daha hızlı bir nöral öğrenmeyi sağladıkları görülmektedir. Bu denemelerde ise EXOR1T ve EXNOR1T olarak işaretlenen eğitim setlerinin optimum x_h değerine yakın ortalama değerlere sahip olmalarından dolayı, EXOR2T, EXOR3T ve EXNOR2T, EXNOR3T olarak işaretlenen eğitim setlerine nazaran daha hızlı bir öğrenme sağladıkları görülmektedir.

5. SONUÇ

YSA topolojisinin, ikili (binary) eğitim veri setinin, saklı katman nöronları için optimum aktivasyon seviyesini nasıl belirlediği ve bu doğrultuda daha hızlı bir nöral eğitim için ikili eğitim veri setinin ortalama değerinin nasıl modife edilmesi gerektiği gösterilmiştir. Elde edilen sonuçlar saklı katman optimum aktivasyon seviyesinin giriş ve çıkış katmanlarındaki nöron sayısına bağımlılığını ortaya koymaktadır. Bu durumda eğitim veri setinin optimum bir sunumunun oluşturulması için eğitim setinin nöral öğrenme üzerindeki etkisi göz önüne alınmalı ve uygun bir kodlama yapılarak ağ topolojisinin problem

öğrenmeyi hızlandırabilecek bir yapıya kavuşturulması gerekmektedir.

TEŞEKKÜR

Bu çalışma Niğde Üniversitesi Araştırma Fonu tarafından F99/04 nolu proje çerçevesinde desteklenmektedir.

KAYNAKLAR

- [1] KODİYALAM, S. and GURUMOORTHY, R., "Neural networks with modified backpropagation learning applied to structural optimisation", *AIAA Journal*, 34, pp. 408-412, 1996
- [2]. H.B. KIM, S.H. JUNG, T.G. KIM, K.H. PARK, "Fast learning-method for backpropagation neural-network by evolutionary adaptation of learning rates". *Neurocomputing*, 11 (1) , pp. 101-106, 1996
- [3]. JEAN, J.S.N. and WANG, J., "Weight smoothing to improve network generalisation," *IEEE Transactions on Neural Networks*, 5 (5), pp.752-763, 1994
- [4]. HASSIBI, B. and STORK, D.G.: 'Second order derivatives for network pruning: optimal brain surgeon.' In *Advances in Neural Information Processing Systems 5*. (Morgan Kaufmann, 1993)
- [5] J Jia and HC Chua, "Neural Network Encoding Approach Comparison: An Empirical Study". *Proc. 1st New Zealand Int. Two Stream Conf. On Artificial Neural Networks and Expert Systems*, pp. 38-41, 1993
- [6] H.C. Chua, "Solving Two-Sprial Problem Through Input Data Encoding", *Electronic Letters*, Vol. 31(10), pp. 813-814, 1995
- [7] H Altun and K. M. Curtis, "Exploiting the Statistical Characteristics of the Speech Signals for Improved Neural Learning in MLP Neural Networks". *IEEE Workshop on Neural Networks for Signal Processing, NNSP'98*, pp. 547-556, Cambridge. UK. August 1998