

YAPAY SİNİR AĞI MİMARİLERİ

ÖZET

Bu makalede 1990 yılından bugüne kadar yapay sinir ağları için donanım geliştirme konusunda çeşitli dergilerde ve konferans bildiri kitaplarında yayınlanmış olan makale ve bildirilerden derlenen bilgiler ışığında önce örneklerle teknolojiye bugün varılan, nokta aktarılmakta, daha sonra yakın ve uzak gelecekte yapay sinir ağları donanımı konusunda ne gibi ilerlemeler görmeyi bekleyeceğimiz konusu kısaca tartışılmaktadır.

IŞIK A YBAY

ODTÜ Elektrik ve Elektronik Mühendisliği Bölümü

1. GİRİŞ

1993 yılı Mayıs ayı ortalarında ODTÜ'de düzenlenen "Hukuksal ve Felsefi Boyutları ile Yapay Zeka" adlı toplantıda düzenleme komitesi benden toplantının başlığı ile pekt c uyuşmayan bir konuda bir konuşma yapmamı istemişti.

Acaba günümüzde yapay zeka ve özellikle yapay sinir ağları konusunda donanım alanında nasıl bir yere gelmiştik? Yani. yapay zeka mimarilerini destekleyebilecek özel amaçlı donanım olarak etrafta neler vardı, kapasiteleri neydi ve nasıl bir yaklaşımı temsil ediyorlardı.

IEEE Transactions on Neural Networks, Transactions on Data and Knowledge Engineering, Transactions on Pattern Analysis and Machine Intelligence, Proceedings of the IEEE dergileri ile 1992 yılı Kasım ayında Çin Halk Cumhuriyeti'nin Beijing şehrinde toplanan International Joint Conference on Neural Networks (IJCNN'92) adlı konferansın bildiriler kitabı ve yapay zeka konusunda çeşitli kitaplarından yararlanarak (bkz. kaynakça) "Yapay Zeka: Teknolojide Nereye Kadar?" başlıklı konuşmamı hazırladım.

Konuşmama yapay zeka konusu ile ilgili genel düşüncelerimi sıralayarak başlamıştım. Ancak değerli meslektaşım Uğur Halıcı benden konuşmamı EMO dergisinin özel bir sayısı için makale haline getirmemi isteyince bu genel düşünceler bölümünün makalede yer almasını uygun görmedim. Bu nedenle

okumakta olduğunuz makale yapay sinir ağları mimarileri konusunda bugün varılan noktayı yeterince aktarabileceğine inandığını örnek mimariler ve konuşmamın sonunda yer alan gelecekte neler göreceğiz' sorusunun tartışıldığı bölümden oluşturuldu.

Ayrıca seçilen örnek mimarilerin tümleşik devre olarak en azından prototiplerinin geliştirilmiş olmasına da dikkat edildi.

Bugüne kadar geliştirilen yapay sinir ağlarına baktığımızda, bunların çoğunun seri bilgisayarlar üzerinde benzetim tekniği ile uyarlanmış yazılımlarından oluştuğunu görüyoruz. Çok çeşitli sinir ağı algoritmaları geliştirilmiş ve bu algoritmalar genellikle seri bilgisayarlar üzerinde yazılım ortamında sınanmıştır.

Sinir ağı yaklaşımı ile geliştirilen algoritmalara baktığımızda, bunların insan beyninin işlevlerini taklit etmek amacı ile yazıldığını görüyoruz.

Bu yaklaşım ile, geleneksel seri bilgisayarlarla çözülmesi çok uzun zaman alabilecek, örneğin görüntü veya ses işleme gibi işlemlerin daha etkin olarak gerçekleştirilebileceği iddia edilmekte.

Son yıllarda yazılım çemberinin dışına çıkarak, bir çeşit sinir ağı bilgisayarı geliştirmeye yönelik çabaların yoğunlaştığı ve birçok tasarımın önerildiği görülüyor.

Bu tasarımlar için genelde izlenen

süreç Şekil 1'de gösterilmiştir.

Sinir ağı bilgisayarı tasarımlarında ortak yaklaşımın basit temel

Oldukça basit yapısı olan İşlemci Birimleri devrenin temelini oluşturmaktadır. İşlemci Birimlerinin (İB) iç yapısı Şekil 2'de



öğelerin çok sayıda tekrarlanması ile oluşan, yoğun ara bağlantılarla desteklenmiş mimarilere yönelme olduğu gözleniyor.

2. ÖRNEK YAPAY SINIR AĞI MİMARİLERİ

Bu bölümde bugüne kadar önerilmiş yapay sinir ağı mimarilerinden önemli bulunan ve donanım tasarımı tamamlanarak prototipleri üretilmiş birkaç örnek sunulacaktır.

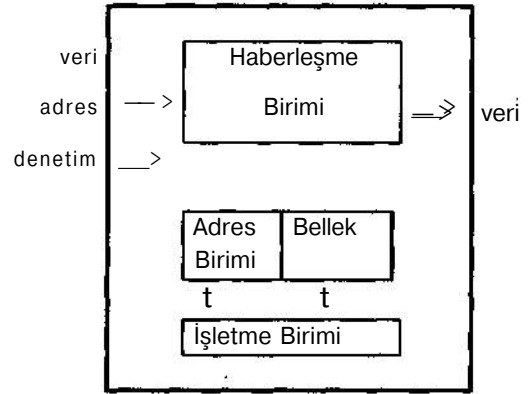
2.1 ELEKTRO NİC NEUROCOMPUTER Vellasco ve Treleaven

Vellasco ve Treleaven, IJCNN'92 de sunulan çalışmalarında ürettikleri bir prototip tümleşik devreyi tanıtmışlardır (1).

Bu tümleşik devre, öğrenmeli Tam Geri Yayılma (Full Back Propagation with Learning) algoritmasını uygulamaktadır.

görülmektedir.

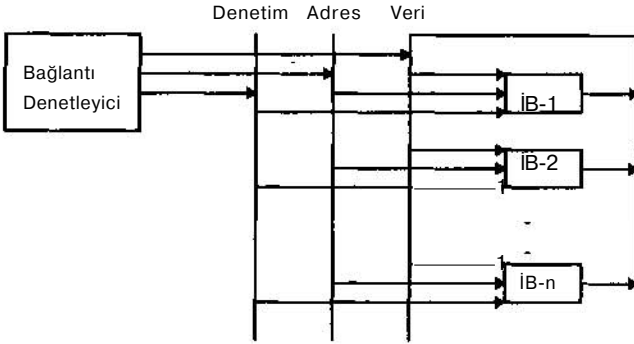
Önerilen mimaride üç değişik bağlantı yer almakta (Denetim, Adres, Veri) ve geniş yayım (broadcast) tekniği kullanılmaktadır. Mimarinin genel yapısı Şekil 3'de gösterilmiştir.



İB'ler içinde yer alan Bellek birimleri sinir ağının durum (state), ağırlık (weight) ve hata (error) bilgilerini tutmaktadır. İşlem Birimlerinde ise aritmetiksel ve mantıksal işlem birimi (ALU), kaydırma (shifter), durum/hata

yazmaçları ve ROM üzerinde değcrasını (threshold) işlevi için bir danışma (lookup) tablosu yer almaktadır.

maktadır. (1) Tek bir yonga üzerinde 16 bitlik bir RISC nükroişlemci. bir haberleşme birimi ve bellekten oluşan temel yapıların herbiri "neural-RISC" olarak adlandırılmaktadır



Mikroişlemcinin % 86'sı tek evrede (single-cycle) işleyebildi 16 komutu vardır. Haberleşme birimi transputer'ler gibi 4 tane seri

Gerçekleştirilen tümleşik devre, 2 um CMOS teknolojisi ile üretilmiştir. Üzerinde sadece 2 İB yer almaktadır. İB'ler 16-bit işlemciler olarak tasarlanmışlardır.

Her İB için 256* 16 bitlik veri RAM bellekleri konmuştur. Her İB içinde 128*16 bitlik bir Danışma Tablosu (Lookup Table) vardır. Tümleşik devre 68 bacaklı olup 76 mm kare yer tutmaktadır. Üzerinde 433 standart ve 179 özel tasarlanmış (ful-custom) hücre (celi) bulunmaktadır.

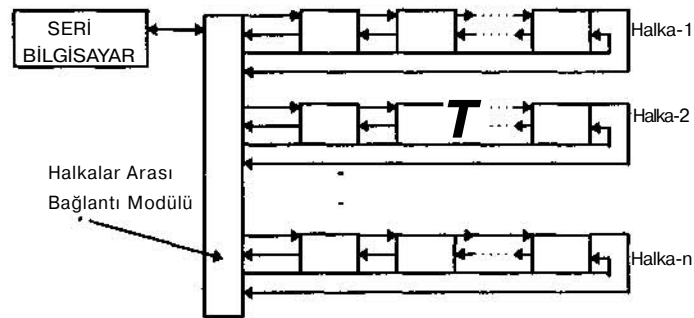
2.2 NEURAL-RISC, Pacheco ve Treleaven

Pacheco ve Treleaven, gene IJCNN'92 de yayınlanan çalışmalarında. 65536 işlemciden (nörondan) oluşacak genel amaçlı bir "Sinir Ağı Bilgisayar"ı için geliştirdikleri transputer benzeri bir RISC (reduced instruction set computer) mikroişlemcisini anlat-

noktadan-noktaya bağlantı ile bilgi paketlerinin dağıtımı için basit bir protokol içermektedir.

Temel yapı taşları olan neural-RISC'lerden bir bilgisayar oluşturabilmek için önerilen "neural-RISC array" adlı bağlantı yapısı ise şekil 4'de gösterilmektedir.

Pacheco ve Treleaven tarafından



gerçekleştirilen bu devrenin 2 um CMOS teknolojisi ile üretilen prototipinin bazı özellikleri şöyle

sıralanabilir:

- Yaklaşık 60.000 transistor.
- 84 mm kare yonga alanı.
- 891 standart. 331 özel tasarlanmış hücre.
- 84 bacaklı bir yonga.
- 16-bitlik işlemci.
- 128*16 bitlik komut ve veri belleği (RAM).
- Her işlemci için 2*(32*16 bitlik) Giriş/Çıkış tamponu.
- 32*16 bitlik başlatma (bool) ROM belleği.
- 100 nanosaniye olarak öngörülen döngü /amanı (eyele time).

64 adet her biri yonga başına 1K* 16 bitlik komut/veri bellekli 16 Neural-RISC işlemciden oluşan Neural-RISC yongası ile 1024 işlemcili bir Neuro-Computer'in tek bir SUN VME kartına sığacağı iddin edilmektedir.

2.3 NEURAL COPROCESSOR, Speckmann, Thole ve Rosenstiel

Almanya Tübingen Üniversitesi'ten Specmann ve arkadaşlarınınca

gerçekleştirilen bir tasarımda Koltonen'in Kendi Kendine Organize Olabilen Özellikler

Haritası (Self- Organizing Feature Map) yöntemi ile çalışan bir sinir ağı yardımcı işlemcisi (coprocessor) geliştirilmiştir. Tasarım, sinir ağlarına büyük verilerin öğretilmesinde karşılaşılan güçlükleri hafifletmek amacıyla taşımaktadır (1).

2.4 CONNECTION MACHINE Üzerinde Geri Yayılma Algoritması, Zhang

Zhang'ın çalışmasında. 32.768 işlemcili bir Connection Machine-2 üzerinde geri yayılma algoritması çalıştırılmıştır. Algoritma veri- paralel olarak çalışmakta, aynı katman (layer) üzerindeki işlemciler için işlemler paralel olarak gerçekleştirilmektedir.

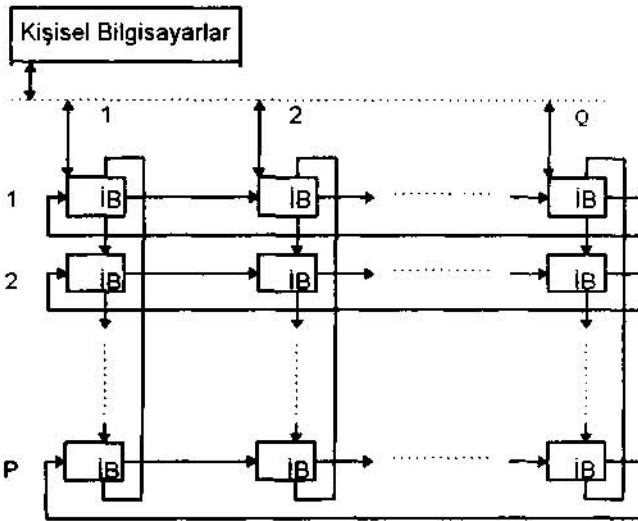
Ayrıca bir anda birden çok örnekle öğretim yapılabilir. Her işlemci bir nöron gibi kullanılmaktadır. Saniyede yapılan ağırlık güncelleme (weight update) 150 milyona ulaşabilmektedir. Bu sayı Pittsburgh Süper Bilgisayar Merkezi'nde 1992 yılında bir Cray sistemi üzerinde ancak 20 milyon olarak elde edilebilmişti (1).

2.5 MASSIVELY PARALLEL NEUROCOMPUTER, Erdoğan, tVahap

Süleyman Şevki Erdoğan ve A. Vahap tarafından Singapur'da yapılan bir çalışmada hem denetim, hem de bağlantılar yönünden tekrar oluşturulmaya açık (reconfigurable) bir geniş ölçekte paralel neurocomputer yapısı önerilmiştir (1). 1992 Yılında, bu teknikle, bir kartta 40 Mhz'lik 12 işlemcili bir prototip yapılmaktaydı. RM-12 adlı bu prototipin beklenen performansı Erdoğan ve Vahap tarafından şöyle duyurulmuştur:

sistem	saniyede yapılan bağlantı yenileme
CRAY-2	13 milyon
Connection Machine	7 milyon
iNVARP	36 milyon
Hitachi	150 milyon
DAP-610	160 milyon
RM-12	200 milyon

Erdoğan ve Vahap. sinir ağı bilgisayarlarının 1 ile 16 karttan oluşacağını söylemektedir. Bir kartın yapısını ise Şekil 5'deki gibi öngörmüşlerdir.



2.6 50 NÖRON LU ANALOG YONGA, Salam ve Wang

1991 Yılında yayınlanan çalışmalarında Salam ve Wang gerçekleştirdikleri krakter ve örii (pallern) tanıyabildi 64 bacaklı bir analog tümleşik devrenin özelliklerini anlatmaktadır (4). 2 tını CMOS teknolojisi ile üretilen tümleşik devrenin alanı 6.8*4.6 mm kare olup, 50 sayısal çıkış kanalını desteklemektedir. Devre, yonga üzerinde sayısal öğrenme (on-chip digital learning) özelliğine sahiptir.

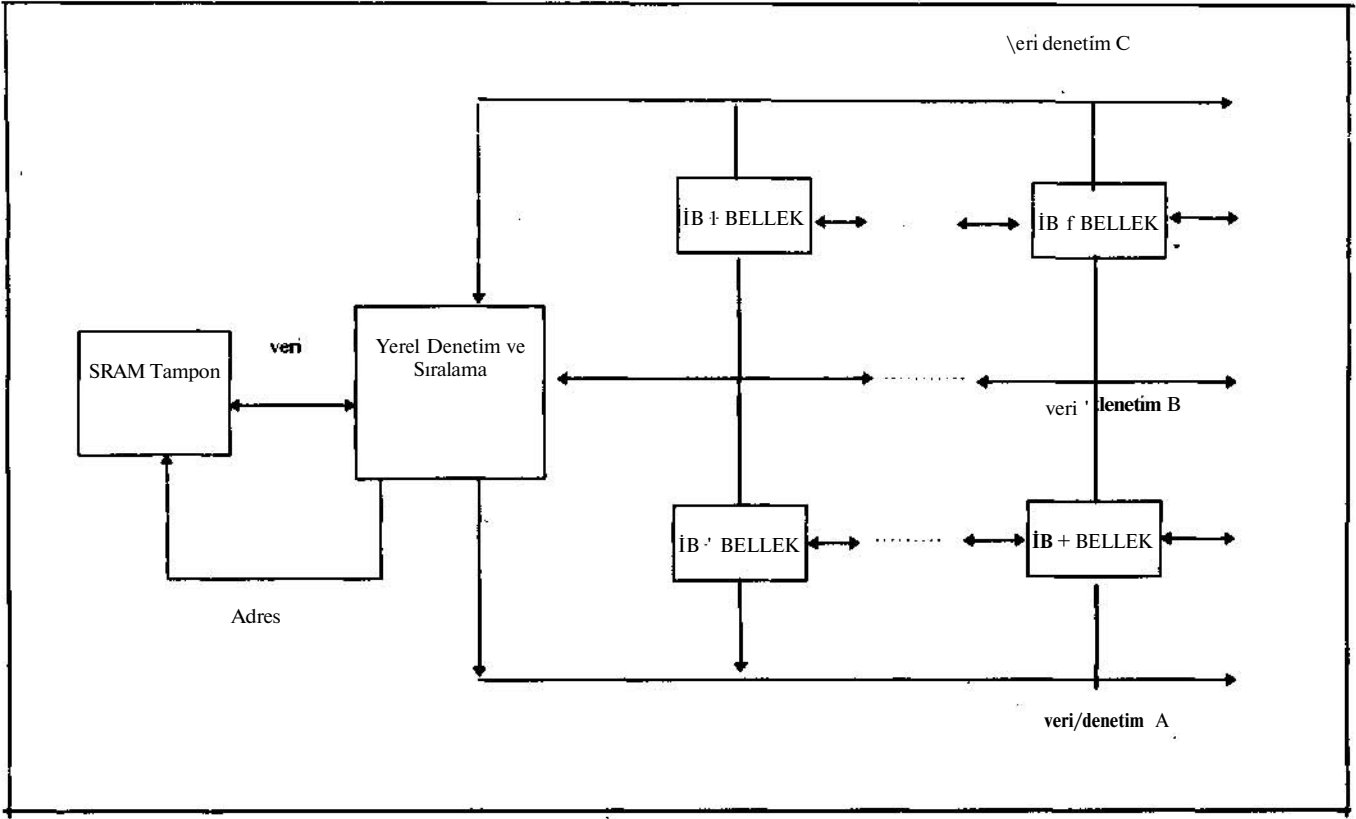
Kullanılan model en çok bağlantı sayısını $n*(n+1)/2$ "yc indirmekle, ara bağlantılarının simetrik olmasını şart koşunamakla. nöronlara kendi kendine geri besleme özelliğini de sağlamaktadır.

Devre üzerindeki 50 nöronun 49 tanesi 7*7'lik bir matris oluşturmakta ve iki boyutlu görüntüleri işleyebilmektedir. Ağırlıklar, yonga üzerinde yer alan öğrenme devreleri yardımı ile programlanabilmektedir.

2.7 YAPAY SİNİR AĞI BENZETİMİ İÇİN GENİŞ ÖLÇEKTE PARALEL MİMARİ, Fujimoto ve Arkadaşları

Japonya'da Sharp sirkeline Fujimoto ve arkadaşlarıncı yapılan çalışmalar sonucunda yapay sinir ağlarının benzetimi için iki adet paralel mimari geliştirilmiştir. Bunlar "Toroidal Lattice Architecture (TLA)" ve "Planar Lattice Architecture (PLA)" olarak adlandırılmıştır (4).

Yapay sinir ağlarının bu mimariler üzerinde benzetimi için düşünülen yöntem, çok katmanlı



perceptron'un geri yayımlı öğrenme algoritması ile "sanal" (virtual) işlemciler üzerinde benzetimi, daha sonra sanal işlemcilerden işin gerçek işlemcilere uyarlanması olarak özetlenebilir.

Bunlardan TLA mimarisi, transputerler kullanılarak gerçekleştirilmiştir. Hopfield yapay sinir ağı ve $256 \times 64 \times 256$ 'lık bir perceptron ile 20 şehre kadar dolaşan satıcı (travelling salesman) problemi çözülmüştür. 16 adet Inmos T800 transputerinden oluşan TLA üzerinde ileri besleme (feedforward) ağında saniyede 2 milyon bağlantıya, geri yayılma ağında ise saniyede 600.000 bağlantı yenilemeye çıkılabilmektedir. Kullanılan TLA mimarisi şekil 6'da gösterilmiştir.

Fujimoto ve arkadaşları, performansın işlemci sayısı ile doğru

orantılı olarak arttığını söylemektedirler. İşlemci Birimlerinde 80 MFLOPS İ860 RISC işlemcileri ve her bir işlemcide 2MB bellek kullanarak gerçekleştirilecek 256×256 matris TLA yapısının

saniyede 100 milyar bağlantı yenileme sınırına kadar çıkabileceğini iddia etmektedirler.

3. GELECEKTEN NELER BEKLEYEBİLİRİZ?



Genellikle daha zeki olan bir insanın daha az zeki olan insanlara göre daha az bilgi ile başarılı olabileceği düşünülür. Ama Simon'a göre (2) insanlar iyi bildikleri konuların dışına çıkarılsa çıkarımlarının ve sonuçlarının (inferences) kuşkulu hale geldiği görülmektedir. Demek ki, başarılı olabilmek için hem güçlü bir düşünce hem de bilgiye ihtiyaç vardır. Yapay sinir ağı mimarilerinde güçlü düşünce yeteneğini birimler, nöronlar arası bağlantıların, bilgi miktarını da bellek kapasitesinin temsil ettiği düşünülürse, geleceğin yapay sinir ağı mimarilerinde hem daha yoğun ara bağlantılarına, hem de daha büyük belleklere gereksinimimiz olacaktır.

Seri-paralel mimari tartışmasında paralel mimarilerin hakim olacağını düşünenlerin sayısı çok fazladır. Günümüzde geliştirilen mimarilerin çoğu genel değil, özel amaçlı paralel mimarilerdir. Ayrıca "detaylarda kaybolmamak" için, belli temel işlevleri gören birbirine yakın işlemcilerden oluşan gruplar (clusters) ve bunların hiyerarşik olarak bağlanması ile daha karmaşık işleri gerçekleştiren mimarilerin çoğalması beklenebilir.

Gelelim neden insanlara birçok şeyi öğrettiğimiz gibi bilgisayarlara da öğretemiyelim sorusuna. Simon'ın da değindiği gibi insanları "programlamıyoruz". 1-2 yaşlarından 20 yaşlarına kadar süren bir öğrenme sürecinde çeşitli bilgileri öğretiyoruz. Bu öğrenme sürecinde insan sürekli olarak dağarcığına yeni bilgileri ekliyor ve çıkarımlar ve sonuçlar için belli bir olgunluk

düzeyine ulaşıyor. İşte bu öğrenme süreci çözümlenip tüm ayrıntıları ile modellenmeden öğrenebildi bilgisayarlar daha bir süre hayal dünyamızda saklanmaya devam edecekler.

Ben bellek kapasitesi konusunda 2000 yılı civarında insan belleğine eşdeğer bir noktaya ulaşabileceğimizi düşünüyorum. Dağınık (distributed) bilgi işleme giderek daha çok önem kazanacak. Belli alanlarda belli büyüklükteki problemleri çözebilen, bir yerde belki o kısıtlı alan içinde bir çeşit "düşünebilen" bilgisayarların sayısı artacak. Ancak işlemciler ve nöronlar arası bağlantı karmaşıklığında insan beyni düzeyine ulaşma yolunda daha uzun bir yolumuz olduğu kanısındayım.

• t. KAYNAKÇA

1. International Joint Conference on Neural Networks, Beijing, China. November 1992, Bildiriler Kitabı. Cilt-3. Electronic Neucomputer

bölümü, s. 165-344.

2. IEEE Transactions on Knowledge and Data Engineering. June 1991. Vol:3. No:2. Yapay Zeka ve Yönetim Özel Sayısı.
3. IEEE Transactions on Neural Networks. Neural Network Hardware özel sayıları. Mart 1991 ve Mayıs 1992.
4. IEEE Transactions on Neural Networks. July 1991 ve November 1992 sayıları.
5. Proceedings of the IEEE. April 1991. Massively Parallel Computers özel sayısı.
6. IEEE Transactions on Pattern Analysis and Machine Intelligence. May 1991. Vol:13. No:5. "The Warp Machine on Navlab". Crispan ve Webb. s. 451-465.
7. Winston. Patrick Henry. "Artificial Intelligence" 2. baskı. Addison Wesley. 1984.

