Ses İşaretlerinin Yapay Sinir Ağları ile Tanınması ve Kontrol İşlemleri için Kullanılması

İpek Barış, Meltem Erdamar, Emre Sümer, Hamit Erdem
Başkent Üniversitesi
Elektronik Mühendisliği Bölümü
Ankara
ipek_baris@hotmail.com,98130024@mail.baskent.edu.tr,{esumer,herdem}@baskent.edu.tr

Özet: *Bu çalışmada Yapay Sinir Ağları (YSA) kullanılarak konuşmacı kimliğinin belirlenmesi amaçlanmıştır. Kişilerin farklı ses karakteristiklerine sahip olması ve bu özelliklerin matematiksel işlemlerle ortaya çıkarılması sayesinde ses tanıma sistemleri günlük hayatta sıkça kullanılmaya başlanmıştır.Burada, ses karakteristiklerinin ortaya çıkarılması ve YSA ile konuşmacının belirlenmesi üzerinde durulmuştur. 7 kişiden alınan ses örnekleri bir dizi matematiksel işlemlerden geçirilerek her örneğe ait 260 tane cepstral katsayı hesaplanmış ve bunlardan bir veri tabanı oluşturulmuştur. Daha sonra Matlab programında yazılan YSA, oluşturulan veri tabanına göre eğitilip test edilmiştir. YSA'nın çıkışında konuşmacıların kimlikleri %90'lık doğruluk payıyla bulunmaktadır.*

1.Giriş

Ticari ve finansal işlemlerin elektronik sistemlerle yapılabilir hale gelmesi, kişilere ait özel bilgilerin istenmeyen şahıslar tarafından elde edilmesi tehdidini doğurmuştur. Bu sorunu aşmak için güvenlik sistemleri gelişen teknolojiyle sürekli olarak yenilenmektedir. Şahıslara ait ses, resim,parmak izi, iris, el yazısı gibi karakteristikler özellikler yardımıyla yeni güvenlik sistemleri kullanılmaktadır. Ses tanıma sistemleri de bunun bir parçasıdır.

2.Ses İşleme

Bütün konuşma sesleri, farklı frekans değerlerine sahip sinüs dalgalarının doğrusal birleşiminden oluşur. İnsan sesinin frekans değerleri 300Hz-3300Hz arasında değişmektedir. Nyquist Teoremine göre ses frekansının iki katı ve daha büyük örnekleme frekansı ile etkin bir örnekleme yapılır [1]. Bu nedenle, ses örnekleri 8kHz'lik örnekleme frekansı ile kaydedilir. Bu örneklerin YSA'ya giriş olarak verilebilmesi için ses sinyalleri üzerinde bazı değişiklikler yapılmalı, sinyaller gürültüden arındırılmalı ve konuşmacıların ses karakteristiklerini oluşturan katsayılar belirlenmelidir. Bu işlemler için bilgi sıkıştırma ve ses özelliklerinin ortaya çıkarılması gibi tekniklerden faydalanılır. Böylece sistemin çalışma hızı ve performansı artar. Ses özelliklerinin ortaya çıkarılması için Şekil-1'deki blok diyagramda gösterilen işlemler sırasıyla yapılır. Şekilde ilkönce ses kaydı yapılır, kaydedilen ses A/D ile sayısal veriye çevrilir.Bu sinyal üzerinde önvurgu işlemi, çerçeveleme ve LPC katsayılarıyla cepstral katsayılar bulunur.

Şekil 1. Ses Özelliklerinin Ortaya çıkarılması

2.1.Önvurgu İşlemi

Bu bölümde ses sinyalinin frekans uzayındaki değerleri düşük dereceli FIR(finite impulse response) bir filtreden geçirilir. Böylece sinyal gürültüden arındırılmış olur ve sadece sinyal karakteristiğini belirleyen kısımlar elde edilir. Sinyalin gereksiz kısımları ve gürültü atılır. Birinci dereceden LPF(alçak geçirgen filtre)'ın transfer fonksiyonu:

$$H(z)=1 / (1-a*z^{-1}) \qquad a=-0.9375 \qquad (1)$$

2.2.Çerçeveleme ve Pencereleme İşlemi

LPF'den geçen ses sinyallerinin her biri çerçevelere bölünür. Bu çerçevelerin tümünün periyodu aynıdır. Çerçeveler belirli bölgelerde kesişirler. Daha sonra hepsi Hamming Windowing denilen bir pencereleme algoritmasından geçer. Böylece çerçevelere bölünmüş ve pencerelenmiş sinyalin karakteristiklerini taşıyan katsayıları hesaplamak kolaylaşır ve sürekli bir sinyal elde edilir[6].

$$\mathrm{Wn} = \begin{cases} 0.54 - 0.46 * \cos(2 * \Pi * n/(N-1)) & 0 \le n < N \\ 0 & \textit{yada} \end{cases} \tag{2}$$

2.3.Doğrusal Tahmini Kodlama (LPC) İşlemi

Sayısal işaret işleme alanında sesi tanımak için bir kaç algoritma kullanılabilir. Bunların içinde en önemlisi LPC dir. LPC'nin kullanım kolaylığı ve hafızada az yer kapsaması en beligin özellikleridir.[2]Bu teknikteki temel ilke ses örneklerinin geçmişteki örneklere bakılarak tahmin edilmesidir[6]. Ses örneğinin, eski örneklerinin doğrusal birleşimi şeklinde olduğu düşünülüp ses sinyalinin karakteristik katsayıları yaklaşık olarak hesaplanır.[3] Elde edilen yaklaşık sonuç ile gerçek değer arasındaki fark yani hata minimuma indirilir.

$$s' = \sum_{i=1}^{p} a_i * s_{n-i} \tag{3}$$

$$E = \sum_{n=0}^{N-1} e_n^2 = \sum_{n=0}^{N-1} \left(s_n - \sum_{k=1}^{p} a_k * s_{n-k} \right)^2 \tag{4}$$

2.4. Cepstral Analiz

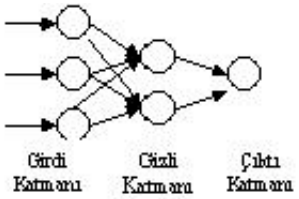
Cepstal analizde bulunan katsayılar LPC katsayılarından türetilir. Bu katsayılar, LPC katsayılarına oranla karakteristik parametrelerin hesaplanmasında daha güvenilir yaklaşımlar sunar. Böylece YSA için verilmesi gereken bilgilerin önemli karakteristik parametreleri elde edilir.[6] Bunlar da YSA'ya giriş olarak verilir.

$$c_k = a_k + 1/k * \sum_{i=1}^{k-1} i * c_i * a_{k-i} \qquad k = 1,2,...,i-1 \qquad 1 < i < p \tag{5}$$

3.Yapay Sinir Ağları ve Ses Tanıma

Yapay sinir ağları (YSA), insan beyninin çalışma prensiplerinin taklit edilmesiyle oluşturulan sistemlerdir. YSA'lar, model seçimi ve sınıflandırılması, fonksiyon tahmini, en uygun değeri bulma ve veri sınıflandırılması gibi işlerde başarılıdır. Geleneksel bilgisayarlarsa özellikle model seçme işinde verimsizdir ve sadece algoritmalı hesaplama işlerinde ve kesin aritmetik işlemlerde hızlıdırlar.[5]

Yapay sinir ağlarında girdiler ve çıktılar arasında gizli katmanlar vardır. Her katmanın girdisi, bir aktivasyon fonksiyonuna girerek çıktıyı oluşturur. Seviyeler arasında ağırlıklı toplamlar ile çıktılar bulunur. YSA'nın giriş-gizli-çıkış katmanları Şekil 2. te görülmektedir. Her nöronun bir ağırlık ve yanlılık değerleri vardır. Yapay sinir ağlarındaki bu ağırlıkları bulmak için değişik algoritmalar kullanılır. İleri besleme – geri yayılım algoritması en yaygın olarak kullanılandır. Bu algoritma ile sisteme sınıflar öğretilir. Böylece sistem eğitilir. YSA'da ses tanımlama işlemleri diğer algoritmalara göre daha hızlıdır. Bu yüzden bu çalışmada YSA kullanılmıştır[7].



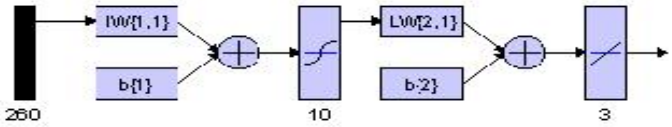
Şekil 2. Yapay Sinir Ağı Modeli

4.Geliştirme

Bu projede, teknolojik gelişmeler paralelinde minimum hata hedefiyle ses sinyallerinin tanınıp konuşmacının belirlenmesi üzerine çalışmalar yapılmıştır. İlk aşamada, 6 kişiden 'a' harfine ait 90 ses tane ses örneği alınmış ve herbir örnekte 260 tane cepstral katsayı bulunan vektörlerle bir veri tabanı oluşturulmuştur. Matlab'da oluşturulan YSA programına giriş olarak veri tabanındaki vektörler verilmiştir. Hedef çıkış değerlerine göre sistem eğitilmiştir.Eğitimde gradyan metodu kullanılmıştır. Eğitim işlemlerinin tamamlanmasının ardından sistem test edilmiştir ve %10'luk hata payı ile konuşmacı doğru olarak tanınmaktadır.

4.1. Yapay Sinir Ağlarında Cepstral Katsayıların İşlemesi

Şekil 3. 'e göre giriş katmanında 260, gizli katmanda 10 ve çıkış katmanında ise 3 nöron kullanılmıştır. Gizli katmandaki nöron sayısının çok olması sistemin etkin çalışmasını sağlar ancak hızı azaltır. Gizli katmandaki optimum nöron sayısı deneme yanılma yoluyla bulunur.

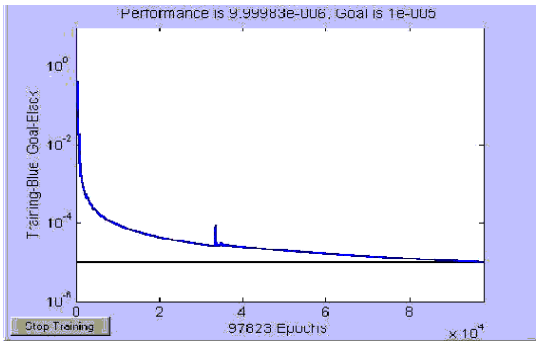


Şekil 3. Tasarlanan Yapay Sinir Ağ Modeli

Çıkış ise tamamen tasarıma bağlıdır . Sistemde konuşmacının sei ve söylediği harf önemlidir. Eğer bu kişiler farklı harf söylerlerse, sistem bu sesi tanımaz dolayısıyla, yabancı birinin sesi olarak algılar. Çıkıştaki 3 nöron aşağıdaki gibi kodlanmıştır. X ve Y farklı kişleri göstermektedir.

Nöron:	Çıkış:		
O ➔	1	0	0
O ➔	0	1	0
O ➔	0	0	1
	X	**Y**	**Yabancı**

Ses örneklerinden oluşturulan vektörler bir hücre dizisi (cell array) altında tutulur.Bu, tasarlanan yapay sinir ağının giriş değeridir. Hedeflenen 3*1'lik vektörler de ayrı bir hücre dizisinde bulunur. Tasarımda gizli katmanın transfer fonksiyonu "tansig", çıkışınki ise "purelin"dir. Sistem "traingd" ile eğitilmektedir. Traingd'nin kullanım amacı, eğitim işleminde tüm verilerin aynı anda kullanılmasıdır.Bu sayede sistemin daha hızlı ve az hata ile çalışması sağlanır. Çünkü, ağırlık değerleri performans fonksiyonunun gradyanının tersi yönde yenilenir. Bunun yanında programda "lr=0.05" öğrenme oranı, "epoch=100000" tekrarlama sayısı, "goal=1*e-005" istenilen tolerans değeri gibi değişkenlere uygun değer atanır ve sistem tüm bu verilerle göre eğitilir. Şekil-4'te YSA'nın performans-amaç eğrisi görülmektedir. Sistemin performans değeri, önceden belirlenen amaç değerine tolerans miktarı kadar yaklaştığı zaman eğitim işlemi sona erer. "perf =mse(e)" ile elde edilen performans değeri : perf = 9.9998e-006



Şekil 4. Tasarlanan Yapay Sinir Ağ Modeline Göre Performans veAmaç Eğrisi

4.2. Test Aşaması

Test amacıyla farklı örnekler sisteme giriş olarak verilir. Çıkış değerleri gözlemlenir.Sonuç ile beklenen değerler karşılaştırılır ve sistemin çalışma performansı hatalara bakılarak hesaplanır.Hata ortalamalı kareler toplamı metoduyla bulunur.

5.Sonuç

Ses örneklerinin alınması, cepstral katsayılar ile veri tabanının oluşturulması, yapay sinir ağının eğitilip test edilmesi sonucunda denenen 10 ses kaydının 9 tanesi doğru olarak bulunmuştur. Bu oranın arttırılması için hem gizli katmandaki nöron sayısı hem de veri tabanındaki ses örneklerinin sayısı arttırılıp sistemin yeniden eğitilmesi gerekir. Ayrıca sistemin çıktıları paralel porta gönderilerek çeşitli cihazların kontrolü de yapılabilir. Bundan sonraki aşamada belirli harfler yerine konuşmacının tüm harflerini tanıyabilen bir sistem tasarlanabilir.

Kaynaklar

[1] Haykin S, Communication Systems. John Wiley, 2000. s: 188-187.
[2]*Botros, N., Deiri, M.Z., Hsu, P.,* Automatic voice recognition using artificial neural network approach Circuits and Systems, 1989.,Proceedings of the 32nd Midwest Symposium on, 1990 s: 763 -765 vol.2
[3] Cansız M., YSA ile Kişilerin Ses Örneklerinden Kimliklerinin Tanınması Yüksek Lisans Tezi, 1997, s:22.
[4] Kuah, K., Bodruzzaman, M., Zein-Sabatto, S, A neural network-based text independent voice recognition system Proceedings of the 1994 IEEE ,1994 . s: 131 -135
[5] Ng, G.S., Erdogan, S.S., Pan, W.N, Neural networks for voice recognition Networks. International Conference on Information Engineering '93. 'Communications and Networks for the Year 2000', Proceedings of IEEE Singapore International Conference on , Volume: 1 , 1993 s: 383 -387 vol.1
[6] Robinson T**.,** Speech Analysis, **http://svr-www.eng.cam.ac.uk/~ajr/SA95,** 1998.
[7] Callan R.,The Essence of Neural Networks,1998. s: 20-56