Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Sigma
2005/2

RECOGNITION OF THE REED INSTRUMENT SOUNDS BY USING STATISTICAL NEURAL NETWORKS

Bülent BOLAT *

Yıldız Teknik Üniversitesi, Elektrik-Elektronik Fak., Elektronik ve Haberleşme Müh. Böl., Yıldız-İSTANBUL

Geliş/Received: 23.08.2004 Kabul/Accepted: 09.03.2005

ABSTRACT

In this paper, reed instrument sounds were recognised by using statistical neural networks. Linear prediction coefficients were used as features. Three different statistical neural network structures were used for this task. The best structure and it's parameters were determined.
Keywords : Sound recognition, computational auditory scene analysis, statistical neural networks, PNN, GRNN, RBF.

KAMIŞLI ENSTRÜMAN SESLERİNİN İSTATİSTİKSEL SİNİR AĞLARI İLE TANINMASI

ÖZET

Bu çalışmada, kamışlı nefesliler ailesinden 4 enstrümanın sesleri, doğrusal öngörü katsayıları kullanılarak istatistiksel sinir ağları yardımıyla sınıflandırılmıştır. Sınıflayıcı olarak 3 farklı istatistiksel sinir ağı kullanılmış ve en yüksek başarıma sahip olan ağ yapısı ve parametreleri tespit edilmiştir.
Anahtar Sözcükler : Ses tanıma, işlemsel işitsel durum analizi, istatistiksel sinir ağları, PNN, GRNN, RBF.

1. GİRİŞ

Müzikal seslerin içerik analizi yapısal kodlama, ses veri tabanı sorgulama sistemleri, otomatik notaya dönüştürme, müzik eğitimi gibi çok çeşitli ve geniş bir uygulama alanına sahiptir. Bu tür uygulamaların önemli alt görevlerinden biri otomatik enstrüman tanıma işlemidir. Enstrüman tanıma işleminin bir başka önemi de, bu alandaki çalışmalardan elde edilen sonuçların konuşma tanıma, konuşmacı tanıma gibi farklı alanlarda da uygulanabilir oluşudur.

19. yüzyılda Helmholtz ile başlayan müzikal sinyallerin tanımlanması problemi, halen kesin bir sonuca ulaşamamıştır. İnsanların ses sinyallerini ne şekilde işledikleri halen çözülememiş problemlerden biridir. Var olan işlemsel işitsel durum analiz (Computational Auditory Scene Analysis - CASA) çalışmaları kabul edilebilir seviyede başarılı olmalarına rağmen halen insan becerilerinin uzağındadır.

Ses sinyallerinin bilgisayarlar yardımıyla tanınması çalışmalarında iki temel işlem basamağı tanımlanmıştır. İlk basamak tanıma işleminde kullanılacak özniteliklerin kestirimi, ikinci basamak ise bu öznitelikleri kullanarak tanıma işlemini gerçekleştirecek bir yapının

e-posta: bbolat@yildiz.edu.tr, tel: (0212) 259 70 70 / 2887

tasarımıdır. Literatürde çeşitli öznitelik grupları tanımlanmıştır. Tanıyıcı yapılar olarak genellikle Gizli Markov Modeller (Hidden Markov Model - HMM), Gauscul Karışım Modelleri (Gaussian Mixture Model - GMM) ve Yapay Sinir Ağları kullanılmaktadır.

Enstrüman seslerinin tanınması yalnızca otomatik sistemler için değil, insanlar için de zor bir görevdir. Genel olarak, enstrüman sayısının arttırılması başarımı hızla düşürmektedir. Brown ve diğerleri [1] bir grup dinleyici ile yaptığı çalışmanın sonucunda 4 enstrümanın ortalama %85 başarım ile doğru sınıflandırıldığını bulmuştur. Bu çalışmada kullanılan enstrüman sesleri obua, klarnet, saksafon ve flüte aittir. Sesler izole notalar şeklinde kaydedilmiştir. Dinleyici grubu, 15 müzisyenden oluşmaktadır. Dinletilen seslerin enstrümanlara göre dağılımı dinleyicilerden gizlenmiştir. Bu deneyde flüt %93 ile en yüksek tanınma oranına erişirken, klarnetin tanınma oranı %71'de kalmıştır. Brown'un bir başka çalışmasında [2] 2 enstrüman (obua ve saksafon) kullanılmış ve %89 başarım elde edilmiştir.

Bu alandaki en geniş araştırma Martin [3] tarafından yapılmıştır. Bu çalışmada 27 enstrüman kullanılmış ve %46 başarıma erişilmiştir. Martin dinleyici olarak 14 müzisyen kullanmıştır. Deney grubuna 14 farklı enstrümanın sesleri sırayla dinletilerek, dinletilen sesin 27 ayrı enstrümandan hangisine ait olabileceği sorulmuştur. Bazı enstrümanlar farklı çalma tekniklerinde çalınmış, bu teknikler de farklı enstrümanlar gibi nitelendirilmiştir. Tek tek enstrümanların tanınma başarımı %46 olarak bulunurken, enstrümanların ait oldukları aileler ise %92 başarımla doğru tespit edilmiştir.

Campbell ve Heller [4] 6 farklı enstrümanla yaptığı bir deneyde %72 başarıma erişirken, Berger'in [5] eski sayılabilecek bir çalışmasında 10 enstrüman için %59 başarım rapor edilmiştir.

Enstrüman seslerinin otomatik tanınması üzerine geniş bir literatür mevcuttur. Bu çalışmalarda çoğunlukla bir işitsel model çıkışı SOM (Self Organising Feature Maps), HMM veya GMM ile sınıflandırılmıştır. Otomatik tanıma sistemleri içinde en fazla enstrümanla yapılan çalışmalar Fujinaga ve MacMillan [6] ile Fraser ve Fujinaga'ya [7] aittir. Fujinaga ve MacMillan 23 enstrümanla yaptıkları çalışmada %68 doğru sınıflama oranına ulaşmıştır. Fraser ve Fujinaga ise 23 enstrümanla %64 başarım elde etmiştir. Bu iki çalışmada öznitelikler spektumdan hesaplanmış ve en iyi sonucu veren öznitelikler bir genetik algoritma ile bulunmuştur. Kaminskyj ve Materka [8] gitar, piyano, marimba ve akordiyonla yaptığı çalışmada %98 doğru sınıflama oranına erişirken, Kostek ve Czyzewski [9] obua, trompet, keman ve çello ile %93 başarım elde etmiştir. Kaminskyj ve Materka öznitelik olarak kareköksel ortalama enerjiyi (RMS) kullanmıştır. Bu çalışmalarda sınıflandırıcı olarak k-en yakın komşuluk ağı (k-NN) kullanılmıştır. Kostek ve Czyzewski ise öznitelikleri spektral zarf ve atak bölgesinden çıkarmış, sınıflandırıcı olarak da 2 katmanlı ileri beslemeli ağ kullanmıştır. Martin,[10] 14 enstrümanla yaptığı çalışmada %72 başarım ile insan sınıflandırıcılardan daha yüksek bir başarım elde etmiştir. Bu çalışmada önce enstrüman ailelerini bulan hiyerarşik bir sınıflandırıcı ve daha sonra enstrümanı tespit eden bir k-NN kullanılmıştır. Benzer bir çalışmada, Eronen [11] 30 enstrümanı %80.6 başarımla sınıflamıştır.

Bu çalışmada kamışlı nefesliler ailesinden 4 enstrümana (bas klarnet, bason, kontrabason ve obua) ait sesler LP (Doğrusal Öngörü– Linear Prediction) katsayıları kullanılarak istatistiksel sinir ağları ile sınıflandırılmıştır. Elde edilen sonuçlar geçmiş çalışmalarla da karşılaştırılarak önerilen ağların başarımı irdelenmiştir.

2. DOĞRUSAL ÖNGÖRÜ

Doğrusal öngörü (LP) analizi ses spektrumunu elde etmenin bir başka yoludur. Burada spektrum, spektral tepeler üzerine yoğunlaşan bir tüm-kutup fonksiyon ile modellenir. LP genellikle konuşma işaretinin analizinde kullanışlı bir yöntem olsa da, müzikal seslerin tanınmasında da kullanılabilir [11]. LP analizi enstrüman tanımada ilk olarak Schmid [12] tarafından kullanılmıştır.

İleri yönlü doğrusal öngörüde amaç, işaretin bir sonraki örneği $\hat{y}$ (n)'i, p adet geçmiş örneğin doğrusal kombinasyonu ile elde etmektir:

$$\hat{y}(n) = \sum_{i=1}^{p} a_i y(n-i) \tag{1}$$

Burada a_i ile gösterilen katsayılar öngörü ya da doğrusal öngörü katsayıları olarak adlandırılır. (1) denklemi, transfer fonksiyonu aşağıdaki gibi olan bir tüm-kutup süzgeç tanımlar:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{2}$$

Doğrusal öngörüde amaç ortalama karesel öngörü hatası e'yi en az yapan a katsayılarını bulmaktır:

$$e = E\left\{ \left| y(n) - \sum_{i=1}^{p} a_i y(n-i) \right|^2 \right\} \approx \sum_{n=-\infty}^{\infty} \left| y(n) - \sum_{i=1}^{p} a_i y(n-i) \right| \tag{3}$$

Burada E{.}, beklendik değer operatörüdür. (3) eşitliğini minimize etmek için çeşitli algoritmalar kullanılmaktadır. Bu çalışmada Levinson-Durbin algoritması kullanılmıştır. Bu algoritma çok bilinen bir yöntem olduğundan burada yer verilmemiştir; ancak detaylarına [13] ve [14]'ten ulaşılabilir.

3. İSTATİSTİKSEL SİNİR AĞLARI

3.1. Radyal tabanlı Sinir Ağları

Radyal tabanlı fonksiyon ağlarında (Radial Basis Functions - RBF) temel fikir, bir grup radyal taban fonksiyonu istenen f fonksiyonuna yaklaşacak şekilde ağırlıklandırarak toplamaktan ibarettir [15]. RBF üç katmanlı bir yapıdır. Giriş katmanı giriş vektör uzayı ile, çıkış katmanı da örüntü sınıfları ile ilişkilidir. Böylelikle tüm yapı, gizli katmanın yapısı ve gizli katman ile çıkış katmanı arasındaki ağırlıkların belirlenmesine indirgenir. Gizli katmandaki nöronların aktivasyon fonksiyonları bir C_j merkezi ve σ_j bant genişliği ile belirlenir. Aktivasyon fonksiyonu,

$$\varphi_j(X) = \exp\left(-\frac{\left\| X - C_j \right\|^2}{2\sigma_j^{\,2}} \right) \tag{4}$$

eşitliği ile tanımlanan bir Gauss eğrisidir. Çıkış katmanındaki j. nöronun çıkışı için genel eşitlik ise şu şekildedir:

$$s_j(X) = \sum_{i=1}^{K} w_{ij} \varphi_i(X) + b_j \tag{5}$$

Burada ω_{ij} gizli nöron i ve çıkış nöronu j arasındaki ağırlık katsayısıdır [16].

3.2. Genel Regresyon Sinir Ağları

Genel regresyon sinir ağları (Generalised Regression Neural Networks - GRNN) RBF ağlarının merkez ve bant genişliklerinin eğitme verisinin deterministik fonksiyonları olarak belirlendiği

özel bir durumudur. Bu sebeple, bu tür ağlarda eğitim tek adımda gerçeklenir ve iteratif yöntemler kullanılmaz.

GRNN ağlarda bir x_i eğitme girişi, ağdaki Gauss çekirdeklerden birinin merkezi olarak atanır. Herhangi bir giriş vektörü x için i. RBF biriminin çıkışı şu şekilde hesaplanır:

$$\beta_i = \exp\left[-\frac{(x-x_i)^T(x-x_i)}{2\sigma^2}\right] \tag{6}$$

Burada σ kullanıcı tarafından belirlenen yumuşatma parametresidir. Herhangi bir x girişi için ağın çıkışı y (7) eşitliği ile hesaplanır.

$$y = \sum_{i=1}^{K} \alpha_i y_i \tag{7}$$

α katsayıları ise şu şekilde hesaplanır:

$$\alpha_i = \frac{\beta_i}{\sum_{i=i}^{K} \beta_i} \tag{8}$$

Eğer giriş vektörü x, herhangi bir x_i eğitme vektörüne yakın ise, x_i'ye ilişkin α_i en büyük olacak ve istenen çıkış y, x_i'ye ilişkin y_i çıkışına yaklaşacaktır [17].

3.3. Olasılıksal Sinir Ağları

Olasılıksal sinir ağları (Probabilistic Neural Network - PNN) Bayes-Parzen kestiriciler olarak da bilinir. K_1 ve K_2 sınıflarından birine ait, m-boyutlu bir x vektörü olsun. K_1 ve K_2 sınıflarına ait olasılık yoğunluk fonksiyonları $F_1(x)$ ve $F_2(x)$ olsun. Bayes Teoremi'ne göre x vektörü,

$$\frac{F_1(x)}{F_2(x)} > \frac{L_1 P_2}{L_2 P_1} \tag{9}$$

eşitsizliği doğru ise K_1, eşitsizliğin tersi doğru ise K_2 sınıfına aittir. Burada P_1 ve P_2, K_1 ve K_2 sınıflarının görülme olasılığıdır. L_1, x vektörünün K_1 sınıfına ait iken K_2 olarak yanlış sınıflama oranı; L_2 ise x vektörünün K_2 sınıfına ait iken K_1 olarak yanlış sınıflama oranıdır ve maliyet fonksiyonu olarak adlandırılır. Buradan görüleceği gibi, $F_1(x)$, $F_2(x)$, L_1 ve L_2'nin bilinmesi durumunda x vektörünün en yüksek olasılıkla hangi sınıfa ait olduğu tespit edilebilir [18]. PNN'lerde sınıflara ait yoğunluk fonksiyonları Parzen pencereleri [19] kullanılarak aşağıdaki şekilde bulunur:

$$F(x) = \frac{1}{(2\pi)^{m/2}\sigma^m n} \sum_{i=1}^{n} \exp\left[-\frac{(x-x_i)^T(x-x_i)}{2\sigma^2}\right] \tag{10}$$

Burada n eğitim verisi sayısı, m giriş uzayının boyutu, i örüntü numarası ve σ ise ayarlanabilir bir yumuşatma terimidir.

4. DENEYSEL ÇALIŞMA

Uygulamada kullanılan ses örnekleri McGill University Master CD Samples veri setinden alınmıştır. Ses örnekleri 44100 Hz örnekleme frekansında ve 16 bit çözünürlükle kaydedilmiştir. Her bir notaya ait ses dosyası 20 ms uzunluklu çerçevelere bölünmüştür. Çerçeveler arasında örtüşme yoktur. Her çerçeve için 6 adet LP katsayısı hesaplanmıştır. Bulunan katsayılar ilk katsayıya göre normalize olduğundan hesaplanan 6 katsayının ilki hep 1 değerine sahiptir. Bu nedenle ilk katsayı atılmış ve her bir çerçeve 5 LPC katsayısından oluşan bir vektörle ifade

edilmiştir. Her bir notanın tüm çerçeveleri için 5 katsayıdan oluşan vektörler hesaplandığında, bu vektörlerin ortalamaları alınarak her bir nota 5 parametre ile temsil edilmiştir.

Çalışmada kullanılan enstrümanların üretebildiği tüm notalar 121 adettir. Hazırlanan veri kümesi her bir enstrümana ait notaları eksiksiz olarak içermektedir. Her bir nota, bir önceki paragrafta belirtilen şekilde elde edilen 5 ortalama LPC katsayısı ile temsil edilmiş ve ağların girişlerine bu katsayılar uygulanmıştır. Her bir enstrümana ait veriler ayrı bir sınıf olarak tanımlanmış, böylelikle problem her bir elemanı 5 parametre ile temsil edilen, 4 sınıflı bir veri kümesinin sınıflandırılması problemine indirgenmiştir. Veri kümesindeki her sınıfın rasgele seçilen %25'i test için ayrılmış, kalan veri eğitme için kullanılmıştır. Kullanılan yapay sinir ağları eğitme verisi ile eğitilmiş, daha sonra test verisi ile test edilmiştir. İstatistiksel ağlarda eğitme işleminde dışarıdan belirlenebilen tek parametre olan bant genişliklerinin en uygun değerleri deneme yanılma yoluyla RBF için 0.05, GRNN için 0.02 ve PNN için 0.01 olarak bulunmuştur. Belirlenen bant genişlikleri kullanılarak yapılan eğitme ve test sonuçları Çizelge 1'de yer almaktadır. En yüksek başarımı veren PNN ağı için her bir enstrümana ait doğru sınıflandırma yüzdeleri Çizelge 2'de gösterilmiştir.

Çizelge 1. Ağların Doğru Sınıflama Oranları

	Eğitme (%)	Test (%)
PNN	100	93.3
GRNN	97.8	90
RBF	100	46.7

Çizelge 2. PNN İçin Doğru Sınıflandırma Yüzdeleri

	Eğitme (%)	Test (%)	Toplam (%)
Obua	100	100	100
Bas Klarnet	100	83.33	96
Bason	100	100	100
Kontra Bason	100	87.5	96.88

5. SONUÇLAR

Yapılan uygulama sonucunda istatistiksel ağlar arasında en yüksek test başarımı PNN ile elde edilmiştir. GRNN %90 test başarımına ulaşırken, RBF'in başarımı %46.4 gibi düşük bir seviyede kalmıştır. RBF ağ eğitme kümesini ezberlemiş, test kümesini kabul edilebilir seviyede doğru sınıflayacak kadar genelleştirme yapamamıştır. RBF ağın hata toleransı yükseltilerek ezberlemesi engellenmeye çalışılmış, ancak bu durumda test başarımının da düştüğü gözlenmiştir. Buradan, RBF ağların bu tür bir uygulama için yetersiz kaldığı söylenebilir. PNN'in toplam başarımı %98.35'tir. Obua ve bason PNN tarafından tamamen doğru sınıflandırılırken, 1 kontrabason örneği bason olarak, 1 bas klarnet örneği de obua olarak yanlış sınıflandırılmıştır. Bu sonuçlara bakarak, PNN ağların enstrüman tanıma işlemlerinde diğer istatistiksel ağlardan daha başarılı olduğu söylenebilir.

Enstrüman tanıma çalışmalarında elde edilen sonuçların geçmiş çalışmalarla karşılaştırılması çeşitli zorluklar içermektedir. Aynı enstrüman gruplarını kullanarak yapılmış çalışmalar bulmak genellikle mümkün olmamaktadır. Bu tür çalışmalarda standart olarak kullanılan ses veri tabanlarının belirli bir standardının olmayışı da bir başka zorluk olarak sayılabilir. Farklı sayılarda enstrümanlar kullanılan çalışmaların karşılaştırılması ise, enstrüman sayısının artması ile birlikte genel başarımın hızla düşme eğiliminde olması nedeniyle pratik değildir. Yine de, kabaca bir değerlendirme yapmak gerekirse; PNN kullanılarak yapılan denemede Brown'un [1] müzisyenlerle yaptığı çalışmadan daha yüksek bir doğruluk oranına

ulaşılmıştır. Diğer geçmiş çalışmalara bakılacak olursa, 4 enstrümanla elde edilen en yüksek başarım Kaminskyj ve Materka [8] tarafından %98 olarak elde edilmiştir. Ancak bu çalışmada kullanılan enstrümanların hepsi farklı ailelerdendir. Kaminskyj ve Materka'nın enstrümanları tam skala yerine bir oktav boyunca değerlendirilmiş olması, yani veri kümesini daraltması başarımı yükselten nedenler arasında sayılabilir.

Uygulama sonuçlarına göre, en yüksek toplam başarım PNN ile %98.35 olarak elde edilmiştir. RBF ağının eğitmede %46.7 gibi düşük bir başarım göstermesine bakılarak bu ağın bu tür görevler için uygun olmadığı söylenebilir. PNN bu ve benzeri ses tanıma görevlerinde başarılı bir ağ yapısı olarak öne çıkmıştır. Geçmiş çalışmalarla yapılan karşılaştırma kesin bir dille olmasa da, bu kanıyı güçlendirecek yöndedir. Brown'ın [1] insan dinleyicilerle yaptığı ilk deneyinde klarnet toplam 4 enstrüman arasında %71 olasılıkla, ikinci deneyinde [2] ise obua 2 enstrüman arasında %89 olasılıkla doğru sınıflandırılmıştır. Bu çalışmada kullanılan PNN obuayı %100, klarneti ise %96 olasılıkla doğru sınıflandırarak insan kulağı ile yarışabilecek doğrulukta olduğunu göstermiştir. Tüm bu bilgilerin ışığında, PNN ağların enstrüman sınıflama için uygun ve yüksek başarımlı bir ağ yapısı olduğuna kanaat getirilmiştir.

KAYNAKLAR

[1] Brown, J. C., Houix, O., and McAdams, S., "Feature Dependence in the Automatic Identification Of Musical Woodwind Instruments", J. Acoustical Society of America, 109, 1064 - 1072, 2001.

[2] Brown, J. C., "Computer Identification of Musical Instruments Using Pattern Recognition with Cepstral Coefficients as Features", J. Acoustical Society of America, 105, 1933 - 1941, 1999.

[3] Martin, K. D., "Sound-Source Recognition: A Theory and Computational Model", PhD. Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1999.

[4] Campbell, W. C. and Heller, J. J., "The Contribution of the Legato Transient to Instrument Identification", Proc. Research Symposium on Psychology and Acoustics of Music, 1978, 30-44.

[5] Berger, K. W., "Some Factors in the Recognition of Timbre", J. Audio Engineering Society, 30, 396-406, 1964.

[6] Fujinaga, I. and MacMillan, K., "Realtime Recognition of Orchestral Instruments", Proc. International Computer Music Conference, Berlin, Germany, 2000, 141-143.

[7] Fraser, A. and Fujinaga, I., "Toward Realtime Recognition of Acoustic Musical Instruments", Proc. International Computer Music Conference, Beijing, China, 1999 175-177.

[8] Kaminskyj, I., and Materka, A., "Automatic Source Identification of Monophonic Musical Instrument Sounds", IEEE Int. Conf. Neural Networks 1995, USA, 1995, 189-194.

[9] Kostek, B. and Czyzewski, A., "Automatic Recognition of Musical Istrument Sounds – Further Developments", Proc. 110th Audio Engineering Society Convention, Amsterdam, Netherlands, 2001.

[10] Martin, K. D., "Musical Instrument Recognition : A Pattern Recognition Approach", Presented at 136th Meeting of the Acoustical Society of America, Norfolk, USA, 1998.

[11] Eronen A., "Automatic Musical Instrument Recognition", MsC. Thesis, Tampere University of Technology, Department of Information Technology, 2001.

[12] Schmid, C. E., "Acoustic Pattern Recognition of Musical Instruments", PhD. Thesis, University of Washington, 1977.

[13] Haykin, S., "Adaptive Filter Theory", Prentice Hall, NY, 1991.

[14] Hayes, M. H., "Statistical Digital Signal Processing and Modelling", John Wiley & Sons, NY, 1996.

[15] Verleysen, M. and Hlavackova, K., "An Optimized RBF Network for Approximation of Functions", Proc. European Symposium on Artificial Neural Networks", Brussels, Belgium, 1994, 175-180.

[16] Paredes, V. and Vidal, E., "A Class-Dependent Weighted Dissimilarity Measure for Nearest Neigbor Classification Problem", Pattern Recogntion Letters, 21, 1027 - 1036, 2000.

[17] Wong, H. S., Wu, M., Joyce, R. A., et al., "A Neural Network Approach for Predicting Network Resource Requirements in Video Transmission Systems", Proc. IEEE Pacific RIM Conference on Multimedia, Sydney, Australia, 2000.

[18] Goh, T. C., "Probabilistic Neural Network for Evaluating Seismic Liquefaction Potential", Proc. IEEE Int. Symposium on Intelligent Systems, Varna, Bulgaria, 2002, 16-20.

[19] Parzen, E., "On Estimation of a Probability Density Function and Mode", Annals of Mathematical Statistics, 33, 1065 - 1076, 1962.