

OLASILIKSAL SINIR AĞLARI İLE MÜZİKAL ENSTRÜMAN AİLELERİNİN TANINMASI

Bülent BOLAT¹

Ünal KÜÇÜK²

^{1,2}Elektronik ve Haberleşme Mühendisliği Bölümü

Elektrik-Elektronik Fakültesi

Yıldız Teknik Üniversitesi, 34349, Beşiktaş, İstanbul

¹e-posta: bbolat@yildiz.edu.tr

²e-posta: kunal@yildiz.edu.tr

Anahtar sözcükler: Ses tanıma, Doğrusal öngörü, Olasılıksal sinir ağları

ÖZET

Bu çalışmada çeşitli enstrümanlara ait sesler kullanılarak bir olasılıksal sinir ağı yardımıyla sınıflandırılması üzerinde durulmuştur. Öznitelik olarak farklı derecelerden doğrusal öngörü katsayıları kullanılmıştır. Yapılan uygulama sonunda, elde edilen en iyi sonucun kabul edilebilir düzeyde olduğu gösterilmiştir.

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.

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

Tanıma işlemi bir adımda gerçekleştirilebileceği gibi, birden fazla adımda da gerçekleştirilebilir. Bu çalışmada, hiyerarşik bir yapıyla gerçekleştirilen bir enstrüman tanıma işleminin ilk adımı olan enstrümanın hangi aileye ait olduğunu tespit etme problemi üzerinde durulmuştur. Enstrümanlar yaylı telliler, pizzicato telliler, çift kamışlılar, flütler, piriç nefesliler, klarnetler ve saksofonlar olmak üzere 7 aileye ayrılmıştır. Öznitelik grubu olarak, literatürde yeterli başarıyı sağlayabildiği görülen ve buna karşılık oldukça hızlı bir şekilde hesaplanabilen Doğrusal Öngörü Katsayıları (LPC) kullanılmıştır. Tanıyıcı olarak ise Olasılıksal Sinir Ağları (PNN) kullanılmıştır.

2. ÖNCEKİ ÇALIŞMALAR

Enstrüman seslerinin tanınması yalnızca otomatik sistemler için değil, insanlar için de zor bir görevdir. Genel olarak, enstrüman sayısının artırılması başarıyı hızla düşürmektedir. Brown ve diğerleri [1] 15 müzisyenden oluşan bir grup dinleyici ile yaptığı çalışmanın sonucunda 4 enstrümanın (obua, klarnet, saksofon ve flüt) ortalama %85 başarımla doğru sınıflandırıldığını bulmuştur. Brown'un bir başka çalışmasında [2] 2 enstrüman (obua ve saksofon) kullanılmış ve %89 başarımla 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. Bazı enstrümanlar farklı çalma tekniklerinde çalınmış, ve bunlar da farklı enstrümanlar gibi değerlendirilmiştir. Bu çalışmada tek tek enstrümanların doğru tanıma yüzdesi %46 olarak bulunurken, enstrümanların ait oldukları aileler %92 başarımla doğru tespit edilmiştir. Campbell ve Heller [4] 6 farklı enstrümanla yaptığı bir deneyde %72 başarımla elde edilirken, Berger'in [5] eski sayılabilecek bir çalışmasında 10 enstrüman için %59 başarımla rapor edilmiştir.

Otomatik tanıma sistemleri içinde en fazla sayıda 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 başarımla elde etmiştir. Bu iki çalışmada öznitelikler spektrumdan hesaplanmış ve en iyi sonucu veren öznitelikler bir genetik algoritma ile bulunmuştur. Martin [8] 14 enstrümanla yaptığı çalışmada %72 başarımla insan sınıflandırıcılardan daha yüksek bir başarımla elde etmiştir. Bu çalışmada önce enstrüman ailelerini bulan hiyerarşik bir sınıflandırıcı ve daha sonra enstrümanı tespit eden bir k-NN kullanılmıştır. Bu çalışmada enstrüman aileleri %93 oranda doğru sınıflandırılmıştır. Benzer bir çalışmada, Eronen [9] 30 enstrümanı %80.6 başarımla sınıflarken, enstrüman ailelerini %94,7 başarımla sınıflayabilmiştir.

3. 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 [9]. LP analizi enstrüman tanımada ilk olarak Schmid [10] tarafından kullanılmıştır.

İleri yönlü doğrusal öngöründe amaç, işaretin bir sonraki örneği (n)'i, p adet geçmiş örneğin doğrusal kombinasyonu ile elde etmektir:

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

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:

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

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

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

Burada $E\{\cdot\}$, beklendiği 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 [11] ve [12]'den ulaşılabilir.

4. OLASILIKSAL SINIR AĞI

Olasılıksal Sinir Ağı (Probabilistic Neural Network - PNN) Bayes-Parzen kestiriciler olarak da bilinir. $K1$ ve $K2$ sınıflarından birine ait, m -boyutlu bir x vektörü olsun. $K1$ ve $K2$ sınıflarına ait olasılık yoğunluk fonksiyonları $F_1(x)$ ve $F_2(x)$ olsun. Bayes Teoremi'ne göre x vektörü,

$$\frac{F_1(x)}{F_2(x)} > \frac{L_1 P_2}{L_2 P_1} \quad (4)$$

eşitsizliği doğru ise $K1$, eşitsizliğin tersi doğru ise $K2$ sınıfına aittir. Burada P_1 ve P_2 , $K1$ ve $K2$ sınıflarının görülme olasılığıdır. L_1 , x vektörünün $K1$ sınıfına ait iken $K2$ olarak yanlış sınıflama oranı; L_2 ise x vektörünün $K2$ sınıfına ait iken $K1$ 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 [13]. Uygulamada genellikle maliyet fonksiyonları eşit alındığından, sınıflara ait olasılık yoğunluk fonksiyonlarının hesaplanması yeterlidir. Sınıflara ait yoğunluk fonksiyonları Parzen pencereleri [14] kullanılarak aşağıdaki şekilde bulunur:

$$F(x) = \frac{1}{(2\pi)^{m/2} \sigma^m} \sum_{i=1}^n \exp \left[-\frac{(x-x_i)^T (x-x_i)}{2\sigma^2} \right] \quad (5)$$

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

5. UYGULAMA

Uygulamada kullanılan ses örnekleri McGill University Master CD Samples veri setinden alınmıştır. Ses örnekleri 44100 Hz örnekleme frekansında ve 16 bit çözünürlükle kaydedilmiştir. LP katsayıları tüm nota boyunca 20 ms uzunluklu çerçeveler için hesaplanmıştır. Çerçeveler arasında örtüşme yoktur. Her çerçeve için LP katsayıları hesaplanmış, tüm çerçeveler için hesaplanan katsayıların ortalamaları alınarak her bir nota bu ortalama değerler ile temsil edilmiştir. Her bir nota için 5, 10, 15 ve 20. dereceden LP katsayıları ayrı ayrı hesaplanmış ve 4 farklı veri kümesi oluşturulmuştur. Oluşturulan veri kümeleri PNN ağlarla sınıflandırılarak LP katsayılarının uzunluğunun başarımına etkisi araştırılmıştır.

Hazırlanan veri tabanında 32 enstrümana ait toplam 974 adet veri vardır. Bu verilerin yaklaşık %30'u (290 adet) test için ayrılmış, kalan veri eğitime için kullanılmıştır. Yapay sinir ağı önce eğitime verisi ile eğitilmiş, daha sonra test verisi ile test edilmiştir. Olasılıksal ağlarda eğitime işleminde dışarıdan belirlenebilen tek parametre olan bant genişliklerinin en uygun değerleri deneme yanılma yoluyla bulunmuştur. Belirlenen bant genişlikleri kullanılarak yapılan eğitime ve test sonuçları Tablo 1'de yer almaktadır.

Tablo-1. Farklı LPC dereceleri için doğru sınıflandırma yüzdeleri

	LPC 5	LPC 10	LPC 15	LPC 20
Eğitime	%100	%99,9	%99,7	%99,9
Test	%51,6	%54,3	%61,6	%58,1
Toplam	%84,6	%86,2	%87,6	%87

6. SONUÇ

Yapılan uygulamalar sonunda en yüksek başarımlar %87,6 ile 15. dereceden LP katsayıları ile elde edilmiştir. Önceki çalışmalar ile karşılaştırıldığında bu

başarımın bir miktar düşük olduğu görülmektedir. Ancak bu tür uygulamalarda kullanılan enstrüman sayısının artmasıyla birlikte başarımlar da düşmektedir. Bu çalışmada kullanılan enstrüman sayısı geçmişteki çalışmalardan daha fazladır. Ayrıca, kullanılan öznelikler de başarıma doğrudan etki etmektedir. Farklı öznelik grupları kullanıldığı durumda daha yüksek başarımlar elde etmek olasıdır, ancak enstrüman tanımada kullanılan diğer öznelikler hesaplama açısından karmaşıktır. Elde edilen sonuçlara göre, işlemsel yükün önemli olduğu durumlarda başarımdan bir miktar feragat edilerek hesaplanması daha kolay olan LPC katsayılarının kullanılabilirliği görülmüştür.

KAYNAKLAR

- [1] Brown, J. C., Houix, O., McAdams, S., Feature Dependence in the Automatic Identification Of Musical Woodwind Instruments, J. ACOUSTICAL SOCIETY OF AMERICA, Vol 109, pp 1064-1072, 2001
- [2] Brown, J. C., Computer Identification of Musical Instruments Using Pattern Recognition with Cepstral Coefficients as Features, ACOUSTICAL SOCIETY OF AMERICA, Vol 105, pp 1933-1941, 1999.
- [3] Martin, K. D., Sound-Source Recognition: A Theory and Computational Model, PHD. THESIS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 1999.
- [4] Campbell, W. C., Heller, J. J., The Contribution of the Legato Transient to Instrument Identification, PROC. RESEARCH SYMPOSIUM ON PSYCHOLOGY AND ACOUSTICS OF MUSIC, pp 30-44, 1978
- [5] Berger, K. W., Some Factors in the Recognition of Timbre, J. AUDIO ENGINEERING SOCIETY, Vol 30, pp 396-406, 1964.
- [6] Fujinaga, I., MacMillan, K., Realtime Recognition of Orchestral Instruments, PROC. INT. COMPUTER MUSIC CONFERENCE, pp 141-143 Berlin, Germany, 2000.
- [7] Fraser, A., Fujinaga, I., Toward Realtime Recognition of Acoustic Musical Instruments, PROC. INT. COMPUTER MUSIC CONFERENCE pp 175-177, Beijing, China, 1999.
- [8] Martin, K. D., Musical Instrument Recognition : A Pattern Recognition Approach, PRESENTED AT 136TH MEETING OF THE ACOUSTICAL SOCIETY OF AMERICA, Norfolk, USA, 1998.
- [9] Eronen A., Automatic Musical Instrument Recognition, MASTER'S THESIS, TAMPERE UNIVERSITY OF TECHNOLOGY, 2001.
- [10] Schmid, C. E., Acoustic Pattern Recognition of Musical Instruments, PHD. THESIS, UNIVERSITY OF WASHINGTON, 1977.
- [11] Haykin, S., Adaptive Filter Theory, PRENTICE HALL, NY, 1991.
- [12] Hayes, M. H., Statistical Digital Signal Processing and Modelling, JOHN WILEY & SONS, NY, 1996.
- [13] Goh, T. C., Probabilistic Neural Network for Evaluating Seismic Liquefaction Potential, PROC. IEEE INT. SYMP.ON INTELLIGENT SYSTEMS, pp 16-20 Varna, Bulgaria, 2002.
- [14] Parzen, E., On Estimation of a Probability Density Function and Mode", ANNALS OF MATHEMATICAL STATISTICS, Vol 33, pp 1065 - 1076, 1962.