

TDA-TABANLI ÇOKSESİL MÜZİK NOTALANDIRMA SVD-BASED POLYPHONIC MUSIC TRANSCRIPTION

Ismail Arı, Umut Şimşekli, Ali Taylan Cemgil, Lale Akarun

Bilgisayar Mühendisliği Bölümü
Boğaziçi Üniversitesi, 34342 Bebek, İstanbul
{ismailar, umut.simsekli, taylan.cemgil, akarun}@boun.edu.tr

ÖZETÇE

Bu çalışmanın amacı çoksesli müzik notalandırılması (transkripsiyon) probleminin verimli bir biçimde çözülmesidir. Problem, doğrusal bir modelleme ile ele alınmış ve çözüm, rassallaştırılmış Tekil Değer Ayrışımı tabanlı yöntem ile hızlandırılmıştır. Yöntemin, literatürdeki en iyi yöntemlerle rekabet edebilecek başarımda çalıştığı gözlemlenmiştir. Büyük verilerin yaygınlaştığı çağımızda geleneksel yöntemlerin yetersiz kaldığı gösterilmiş ve çok büyük verilerde dahi uygulanabilir, rassallaştırılmış matris ayrışımına dayalı bir yöntem ile bu soruna çözüm getirilmiştir. Zaman ve yerden tasarruf sağlanmasına rağmen elde edilen yüksek başarımlar korunmuştur.

ABSTRACT

The aim of this work is to perform polyphonic music transcription in an efficient way. The problem is formulated as a linear model and the speed is improved by a randomized SVD-based method. The method is shown to compete with the best resulting approaches in literature. The conventional methods seem to fail in this era of big data whereas the proposed method efficiently handles this by use of randomized algorithms for matrix decompositions. The method is able to improve time and space complexity without compromising the high success rate.

1. GİRİŞ

Otomatik müzik notalandırma (transkripsiyonu), ses ve müzik işleme araştırmalarında ele alınan en temel problemlerden biridir. Bu problemde amaç, belirli bir ses verisi üzerinde bir veya birden fazla müzik aletiyle aynı anda çalınmış notaları ve bu notaların hangi zamanlarda ve ne kadar süreyle çalındığını kestirebilmektir. Bu problemi çözmekte kullanılan yöntemler fonetik, konuşma işleme ve müziksel bilgi çıkarımı alanları başta olmak üzere birçok alanda önemli bir yere sahiptir [1].

Notalandırma problemi genel olarak perde kestirimi ile ilgili olduğundan, tarihsel süreçte bu problemi çözmek için çoğunlukla perde kestirim yöntemleri uygulanmıştır. Bu yöntemler kabaca algoritmik ve model tabanlı olarak ikiye ayrılabilir. Çapraz ilinti (korelasyon) hesabı algoritmik tabanlı notalandırma yöntemlerinde sıkça kullanılırken, model tabanlı

yöntemler genellikle olasılık modelleri ve ayrışım modelleri üzerine kurulmuştur [2]. Öte yandan, insan işitme sisteminin notaları tam olarak nasıl tanıdığı henüz kesin olarak bilinmemekle birlikte, bu işlemin sadece işaret işlemeye dayalı yöntemlerle gerçekleşmediği düşünülmektedir. Smaragdis bu düşünce çerçevesinde, belirli bir müzik parçasını, daha önceden oluşturulmuş büyük bir veri tabanındaki örneklerle basit bir yöntemle karşılaştırarak başarılı bir notalandırma yöntemi elde edilebileceğini göstermiştir [3].

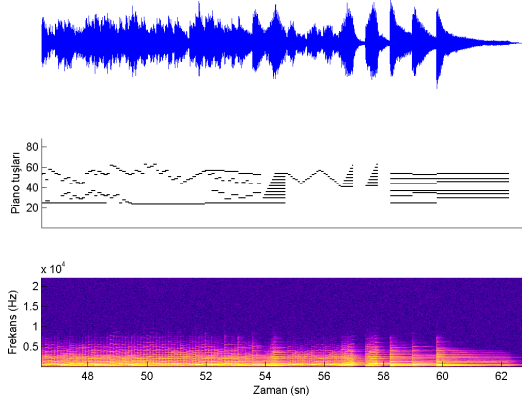
Biz bu çalışmada piyano ile çalınmış çoksesli (polifonik) müzik kayıtlarının otomatik notalandırılması için bir yöntem öneriyoruz. [3]'ü temel alarak önerdiğimiz yöntem, öğrenilmiş veri ile gözlemlenen veri arasındaki ilişkiyi bulmayı hedefleyen ve aynı anda etkin olan notaların seyrekliğini göz önünde bulduran doğrusal bir model kullanarak çıkarım yapmayı hedeflemektedir. Ancak algoritmik olarak maliyetli olan bu yöntem, zaman ve bellekten tasarruf sağlayacak biçimde rassallaştırılmış Tekil Değer Ayrışımı [4] eşliğinde yeniden ele alınıp, başarımlar korunarak daha verimli bir çözüm sunulmaktadır.

2. ÇOKSESİL MÜZİK NOTALANDIRMA

Yöntemi tartışmaya başlamadan önce, yapay öğrenme yoluyla yaklaşacağımız çoksesli müzik notalandırılması probleminin tanımını netleştirelim. Elimizde, $c = 88$ adet notanın çeşitli piyanolarda çalınmış tek sesli (monofonik) kayıtları bulunsun. Bunları $x_i(t)$, $i = 1, \dots, c$ ile gösterelim. Notasyonu kolaylaştırmak için belli bir notaya ait tüm kayıtların birbiri peşi sıra eklendiğini varsayalım ve t zaman indisini göstereyim. Spektrogramlarını, d boyutlu $\mathbf{f}_{x_i}(t)$ ile gösterelim. Her nota için elde edilen örnek vektör sayısı N_i olsun ve toplamda $N = \sum_{i=1}^c N_i$ adet eğitim örneği bulunsun. Örneğin, $\mathbf{f}_{x_{49}}(t)$, piyanodaki 49. tuş olan 440 Hz frekansındaki La kaydının spektrogramındaki t . sütuna karşılık gelmektedir, $t = 1, \dots, N_{49}$. Ses şiddeti konu dahilinde olmadığı için kalan vektörleri, elemanları toplamı 1 olacak şekilde düzgeleyelim. Toplam enerjisi belli bir seviyenin altında olan sütunlar, normalize edilip eğitim kümesine eklenirse ek bir bilgi getirmeyeceği ve gürültüye sebep olacağı için onları eklemeyelim. Bu veri, eğitim kümesini oluşturmaktadır. Test verisi olarak çoksesli kayıtlar kullanılmaktadır. Bu bildiride piyano incelenmektedir, fakat önerilen yöntem birden çok enstrümanın olabileceği genel durum için de geçerlidir. Yani, kayıttaki tek bir enstrümana ait birden fazla nota aynı anda çalınmış olabileceği gibi, farklı enstrümanlar da kullanılmış olabilir. Test kaydını $z(t)$ ile gösterelim ve spektrogramı $\mathbf{f}_z(t)$ ol-

İ. Arı, ve L. Akarun 108E161 nolu, U. Şimşekli ve A. T. Cemgil ise 110E292 nolu araştırma projeleri kapsamında TÜBİTAK tarafından desteklenmektedir. U. Şimşekli'nin çalışması TÜBİTAK BİDEB 2211 bursuyla desteklenmektedir.

sun. Örneğin Şekil 1’de Bach’a ait bir eserin yaklaşık 15 sn’lik bir bölümüne ait dalgaformu, piyano tuşları ve spektrogramı verilmiştir. Etkin piyano tuşlarına bakılınca eserdeki çokseslilik açıkça görülmektedir.



Şekil 1: Çokslesli müzik örneği: Dalgiformu (üstte), notalara denk gelen tuş numaraları (ortada) ve ilgili spektrogram (altta).

2.1. En yakın altuzay yaklaşımı

İncelenen eser teksesli olsaydı, notalandırma için veri kümesi içinde $\mathbf{f}_z(t)$ 'ye en yakın komşu bulunarak onun etiketi seçilebilirdi. Fakat çokslesli durumda $z(t) \approx \sum_{i=1}^c (\alpha_i x_i(\tau_i))$ karışımı gözlemlenmektedir. Benzer şekilde $\mathbf{f}_z(t) \approx \sum_{i=1}^c (\alpha_i \mathbf{f}_{x_i}(\tau_i))$ karışımını da ele alabiliriz. Bu ifade fiziksel gerçekliği tam olarak yanıtmassa da birçok müzik uygulamasında yaygın olarak kullanılmaktadır. Yaklaşımı en iyi ifade eden ağırlıklar bulunduğunda α üzerinde basit bir eşikleme ile hangi notaların etkin olduğu saptanabilir. $D[\cdot, \cdot]$ uzaklık fonksiyonu için ağırlıklar, $\{\alpha_i\}_{i=1}^c$, ve tahmini notalar, $\hat{p}_z(t)$, şöyledir:

$$\{\alpha_i\}_{i=1}^c = \arg \min_{\{\tau_i\}_{i=1}^c} D \left[\mathbf{f}_z(t) \left\| \sum_{i=1}^c (\alpha_i \mathbf{f}_{x_i}(\tau_i)) \right\| \right] \quad (1)$$

$$\hat{p}_z(t) \equiv \{i | \alpha_i > \alpha_{\text{eşik}}\}_{i=1}^c \quad (2)$$

Bu problem, en yakın altuzay problemi olarak bilinmektedir. Örneğin, $c = 2$ durumunda en iyi sonuç, verili kaynaklardan geçen doğrunun üstünde bulunan ve test noktasına en yakın olan nokta ile elde edilir. Toplamda $N_1 \times N_2$ adet farklı seçenek olmakta ve uzaklık fonksiyonu her çağrıldığında d boyut işlenmektedir. Bunun ayrıntılı tartışması için [3]'e başvurulabilir. $c > 2$ için ise, problem $(c - 1)$ 'lik bir altuzayda en yakın noktayı bulma problemine dönüşmekte ve zaman karmaşıklığı $O(d \times \prod_{i=1}^c N_i)$ olmaktadır. Notaların eğitim kümesinde dengeli olarak dağıldığını varsayarsak bu $O(d(\frac{N}{c})^c)$ olarak yazılabilir. Yer karmaşıklığı ise d boyutlu N adet vektör sakladığımız için $O(dN)$ olmaktadır.

2.2. Doğrusal model yaklaşımı

Nota ve enstruman çeşitliliği arttıkça en yakın altuzay yaklaşımının pratikte uygulanabilir bir yöntem olmadığı açıkça görülmektedir. Bunun yerine daha esnek bir çözüm olan

doğrusal model uygulanabilir. Bu modelde gözlemlenen spektrum değerleri, öğrenim kümesindeki spektrum değerlerinin üstdüşümü (süperpozisyonu) olarak ifade edilir. Bir başka deyişle, tüm öğrenim örneklerini yanyana koyarak elde ettiğimiz $\mathbf{F} := [\mathbf{f}_{x_1}(1, \dots, N_1) \ \mathbf{f}_{x_2}(1, \dots, N_2) \ \dots]$ katalog matrisinin sütunlarının doğrusal birleşimidir:

$$\mathbf{f}_z(t) \approx \mathbf{F}\mathbf{w}(t) \quad (3)$$

Notasyonu sade tutmak adına, bundan sonra t indisini dışarıda bırakacağız. Problem bu biçimde ifade edildiğinde, hedef, $D[\mathbf{f}_z \| \mathbf{F}\mathbf{w}]$ 'yi en düşük yapan \mathbf{w} ağırlık vektörünü bulmaktır. İfadeden \mathbf{w} 'nun l_2 -normu da çıkarılarak, bulunan değerler seyrek olmaya zorlanabilir. Toplamı 1 eden bir vektörde l_2 -normu seyrek durumda yüksek olmaktadır. λ seyreklik katsayısı olarak verildiğinde, hedef, $D[\mathbf{f}_z \| \mathbf{F}\mathbf{w}] - \lambda \|\mathbf{w}\|_2^2$ değerini minimize etmektir. $D[\cdot, \cdot]$ için KL-ırsayı kullanıldığında genel yöntem şöyledir: \mathbf{w} rassal değerlerle ilkelendirilir ve yakınsayana kadar her döngüde alttaki üç işlem tekrarlanır [3]:

$$\mathbf{w} \leftarrow \mathbf{w} \odot \begin{pmatrix} \mathbf{F}^T \mathbf{f}_z \\ \mathbf{F}\mathbf{w} \end{pmatrix} \quad (4)$$

$$\mathbf{w} \leftarrow \mathbf{w} + \lambda \frac{\mathbf{w}^2}{\|\mathbf{w}\|_2^2} \quad (5)$$

$$\mathbf{w} \leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|_1} \quad (6)$$

İlk adımda \mathbf{w} değerleri, Negatif-olmayan Matris Ayrışımı'nın KL-ırsayı çözümüne benzer biçimde tahmin edilmekte; ikinci adımda seyrekliğe zorlanmakta; üçüncü de ise normalize edilmektedir. İlk adımdaki \odot sembolü, eleman eleman çarpım işlemidir (Hadamard). Üs alma ve bölmeler de eleman eleman yapılmaktadır. Ağırlık değerlerine bakıldığında etkin notalar diğerlerinden belirgin bir biçimde yüksek olarak gözlemlenmekte ve eşikleme ile seçilmektedir. Eşikleme öncesi, örnek sayısından bağımsız ortak bir eşik değeri kullanabilmek için aynı notaya ait ağırlıklar birleştirilip c boyutlu bir ağırlık vektörü elde edilmektedir. Ayrıca hatanın azaltılması için bu vektörünün her boyutu, medyan süzgecinden geçirilmektedir.

Algoritma incelendiğinde, hesap yükünün (4)'deki \mathbf{F} ve \mathbf{F}^T 'yi içeren matris-vektör çarpımlarında olduğu görülmektedir. Dolayısıyla yöntem, matris-vektörün çarpımının karmaşıklığı olan $O(dN)$ karmaşıklığındadır. Yer karmaşıklığı ise \mathbf{F} 'deki eleman sayısı, yani $O(dN)$ 'dir.

2.3. TDA ile verimlilik arttırımı

Doğrusal yöntem, en yakın altuzay problemine verimli bir çözüm olarak görülmesine rağmen çok büyük verilerin toplanabildiği günümüzde daha verimli yöntemlere ihtiyaç duyulmaktadır. Katalog matrisinde yüzbinlerce sütun bulunabilir, hatta gerçek uygulamalarda bu katalog belleğe (RAM) sığmayabilir. Amacımız, doğrusal modeli hem zaman hem de yer açısından verimli hale getirmektir. Bunun için, \mathbf{F} 'nin doğrudan kullanımı yerine ayrışımının kullanılması düşünülmüştür. Herhangi bir matrisin k mertebeli en iyi yaklaşımı Tekil Değer Ayrışımı (TDA) ile elde edilir [5]:

$$\arg \min_{\mathbf{B}, \text{mertebe}(\mathbf{B})=k} \|\mathbf{F} - \mathbf{B}\|_F = \mathbf{U}_k \Sigma_k \mathbf{V}_k^T \quad (7)$$

Burada k mertebeyi, \mathbf{U}_k sol tekil matrisi, Σ_k en büyük k tekil değerden oluşan köşegen matrisi ve \mathbf{V}_k sağ tekil matrisi göstermektedir. Problem açısından ayrışımı iki matrise indirmek uygundur: $\mathbf{U}_k \Sigma_k \mathbf{V}_k^T = \mathbf{U}_k \tilde{\mathbf{V}}_k^T$. Bu doğrultuda (4) yeniden ele alınıp alttaki biçimde yazılmıştır:

$$\mathbf{w} \leftarrow \mathbf{w} \odot \left(\tilde{\mathbf{V}}_k \left(\mathbf{U}_k^T \frac{\mathbf{f}_z}{\mathbf{U}_k (\tilde{\mathbf{V}}_k^T \mathbf{w})} \right) \right) \quad (8)$$

Bu yönteme başvurulduğunda TDA'nın başlıbaşına masraflı olduğu gözden kaçmayacaktır. TDA'nın tam çözümü $O(\min\{dN^2, d^2N\})$ karmaşıklığındadır [5]. Bizim durumumuzda $d < N$ olduğu için $O(d^2N)$ olacaktır ki bu da veri büyüdükçe pratik olarak mümkün görünmemektedir. Öte yandan, notalandırma için spektrumdaki tüm frekans bantlarına ihtiyaç olmayabilir, örneğin üst doğuşkanlarının tümünü bilmeden de bir notayı tanıyabiliriz. Bu çalışmada, son yıllarda ciddi gelişmelerin kaydedildiği rassallaştırılmış yöntemlerden yararlanılmakta ve Halko *v.d.*'nin geliştirdiği TDA'nın rassallaştırılmış kısmı çözümü kullanılmaktadır [4]. Bu yöntem, \mathbf{F} 'yi bir dönüşüm matrisi olarak ele alır, rassal noktalar üretirek \mathbf{F} ile dönüşüme uğratar ve gözlemlenen değer uzayının dikleştirilmesine dayanır. $d \times N$ boyutlu bir matrisin k mertebeli kısmı TDA'sının kestirimi $O((d + N)k)$ zaman karmaşıklığında bulunmaktadır. Belleğe sığmayan verikümelelerinde dahi ayrıştırma yapılabilir. Rassallaştırılmış matris ayrışmaları hakkında ayrıntılı bilgi için [6]'ya başvurulabilir. TDA eğitim aşamasında yapılmakta, test aşamasında ise elde edilen \mathbf{U}_k ve $\tilde{\mathbf{V}}_k$ matrisleri kullanılmaktadır. İşlemler (8)'deki sırayla yapıldığında zaman karmaşıklığı $O((d + N)k)$ olarak elde edilir. Yer karmaşıklığı, \mathbf{U}_k ve $\tilde{\mathbf{V}}_k$ 'nin toplam eleman sayısı olan $O((d + N)k)$ 'dir. Bahsedilen üç yöntemin zaman ve yer karmaşıklıkları Tablo 1'de özet olarak verilmiştir. Örneğimizdeki $d = 1025$, $N \approx 115600$ için %80 kadar kazanç sağlanmaktadır.

Tablo 1: Yöntemlerin zaman ve yer karmaşıklıkları

	Zaman	Yer
En yakın altuzay	$d(\prod_{i=1}^c N_i)$	dN
Doğrusal model (tam çözüm)	dN	dN
Doğrusal model (TDA)	$(d + N)k$	$(d + N)k$

3. DENEYLER VE SONUÇLAR

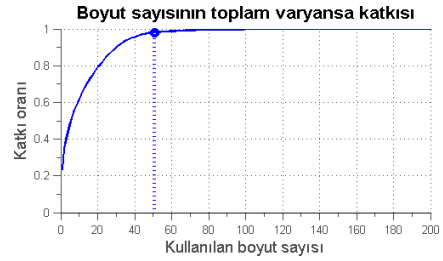
Önerilen yöntemin sınanması için MAPS (MIDI Aligned Piano Sounds) veritabanı kullanılmıştır [7]. 44100 Hz örnekleme frekansındaki 440 adet tek sesli ses dosyası ile eğitim kümesi oluşturulmuştur. Spektrogram verisi, 512 birim atlamalarla elde edilen 2048'lik pencerelerin Hızlı Fourier Dönüşümü'nden (FFT) elde edilmiştir. Hesaplamalarda Hanning penceresi kullanılmıştır. Kullanılan ses dosyalarından elde edilen spektrumun yaklaşık 1/3'ü düşük ses seviyesi sebebiyle eğitim kümesine dahil edilmemiş ve yaklaşık olarak 1025×115600 boyutunda (860 MB'lik) bir katalog matrisi oluşturulmuştur. Test için ise çoksesli 5 farklı eserin çeşitli bölümleri alınarak eğitimdekine benzer biçimde spektrogram elde edilmiştir. Test

kümesindeki çokseslilik dereceleri Şekil 2'de çokseslilik histogramı ile gösterilmiştir.



Şekil 2: Çokseslilik histogramı: Test kümesinde çokseslilik hakimdir ve yoğunluk ilk 6 derecede bulunmaktadır.

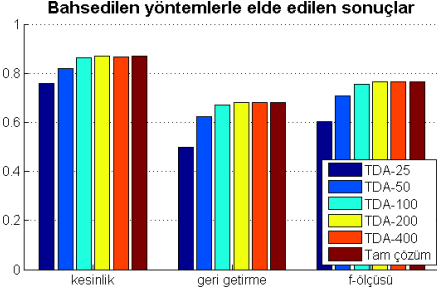
Elde edilen katalog matrisi, rassallaştırılmış TDA yöntemi ile ayrıştırılmış ve özdeğerlerin (tekil teğelerin karelerinin) birikimsel toplamı Şekil 3'de verilmiştir. Çizimden de anlaşılacağı üzere veride yüksek korelasyon bulunmaktadır ve TDA isabetli bir yaklaşımdır.



Şekil 3: Birikimsel özdeğer toplamları: Görüldüğü üzere, veride yüksek korelasyon bulunmaktadır ve toplam varyansın %98'ini ilk 51 tekil vektör içermektedir.

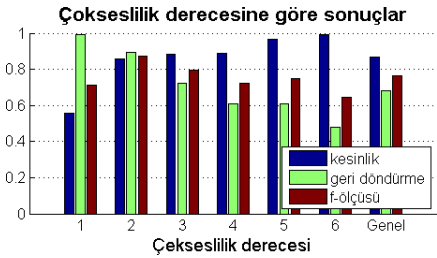
Çalışmamızda k sabit tutulmamış; önerilen yöntem 25, 50, 100, 200 ve 400 değerleri için denenmiştir. Ayrıca karşılaştırma yapmak amacıyla \mathbf{F} 'nin doğrudan kullanıldığı (4)'te verilen tam çözüm de denenmiştir. Başarı ölçütleri olarak kesinlik (*precision*, bulunan notaların doğruluk oranı), geri getirme (*recall*, doğru notaların ne kadarının bulunduğu) ve bunların bileşkesi olan f -ölçüsü $= \frac{2 \times \text{kesinlik} \times \text{geri getirme}}{\text{kesinlik} + \text{geri getirme}}$ oranları kullanılmıştır. En yakın altuzay yaklaşımı pratikte mümkün olmadığı için ona ait sonuç verilmemektedir. Doğrusal yöntemin tam çözümü ve farklı k değerlerine ait ayrışimlarla elde edilen oranlar Şekil 4'de toplu olarak gösterilmiştir. Tam çözüm için f -ölçüsü %76.42 çıkmakta, fakat aynı başarı oranını yakalamak için 200 boyut yeterli olmaktadır. Veri %80 oranda düşürülürken ve hız yaklaşık 5 katına çıkarılırken başarı oranı korunabilmektedir. Eğer spektrogram elde edilirken daha sıkı aralıklarda pencereler alınırsa katalog matrisi dev bir matris olmaktadır. Bu durumda tam çözüm, TDA-tabanlıya göre yavaş kalmakta ve belleğe sığmayacak büyüklükte yer gerektirebilmektedir. Fakat, bir önceki örnekteki benzer sayıda tekil vektör bu matrisi ifade etmek için yeterli olmaktadır.

Yöntemlerin ve seçilen tekil değer sayısının irdelenmesine ek olarak 200'lük TDA-tabanlı yöntem için çokseslilik dere-



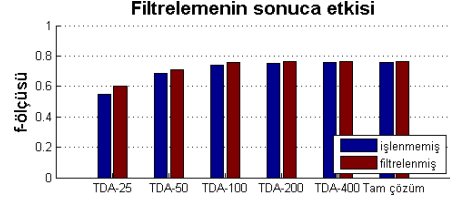
Şekil 4: Test kümesinde elde edilen sonuçlar: Görüldüğü üzere, TDA'da kullanılan boyut sayısı arttıkça başarımlar da doğal olarak artmaktadır. Bununla birlikte, 200 boyut, tam çözümle çok yakın sonuçlar üretmektedir ve yeterlidir.

cesi bazında sonuçlar Şekil 5'de verilmiştir. Elde edilen %76.42 f-ölçüsü literatürde aynı veritabanından elde edilen en iyi f-ölçüsü sonuçlarıyla (%81 [7], %77 [2]) rekabet edebilecek seviyededir. Yöntemin öğrenmeye dayalı sade bir yöntem olduğu ve alana özel ileri sinyal işleme teknikleri kullanmadığı unutulmamalıdır. Şekilde görüldüğü üzere, teksesli durumda geri getirme mükemmel çalışırken kesinlik değeri düşük görünmektedir. Seçimde w üstünde kullandığımız eşik değerini arttırsak daha az notayı kabul eder ve kesinlik değerinde artış elde ederiz. Fakat geri getirme düşecektir. Şekildeki durumda, çokseslilik derecesi arttıkça bulunan notaların daha kesin olduğu fakat olması gereken bazı notaların kaçırıldığını görmekteyiz. Kesinlik ve geri getirmenin bileşkesi olan f-ölçüsünde en yüksek değeri elde etmek için eşik değeri ve λ üstünde uygulamaya özel eniyilemeye gidilebilir. Örneğin verideki çokseslilik hakkında önbilgimiz varsa λ yüksek tutularak sonuç daha seyrek olmaya zorlanabilir.



Şekil 5: Farklı çokseslilik derecelerine göre kesinlik, geri getirme ve f-ölçüsü değerleri: Şekil, 200 boyutlu TDA kullanılarak üretilmiştir. Tam çözüm ile yakın sonuçlar verdiği için onun için ayrıca sonuç eklenmemiştir. Şekilde, f-ölçüsünün her derece için %65'ten yüksek olduğu görülmektedir.

Raporlanan sonuçlar elde edilirken kullanılan λ ve eşik değerleri, test kümesi haricinde bir geçiş kümesi üstünde eniyileme ile elde edilmiştir. Ayrıca, ardıl-işlem olarak medyan süzgeci ile filtreleme yapılmıştır. Ağırlık değerlerine süzgeç uygulanmadan ve uygulanarak eşikleme yapıldığında elde edilen f-ölçüsü oranları Şekil 6'da görülmektedir. Filtreleme $w(t)$ 'nin her elemanı için, t zaman doğrultusunda 15'lik medyan süzgeci ile (eşiklemeden önce) yapılmıştır.



Şekil 6: Medyan süzgeci kullanımının f-ölçüsüne etkisi: Tam çözümde ve TDA-tabanlı çözümün her durumunda filtrelemenin olumlu katkısının olduğu görülmektedir.

4. VARGILAR

Bu bildiriye çoksesli müzik notalandırılması problemi irdelenmiş ve çokseslilik derecesi arttıkça en yakın altuzay yaklaşımının pratikte uygulanabilir bir yöntem olmadığı gösterilmiştir. Problem, doğrusal modellemeyle yeniden ele alınmış, eğitim kümesinin çok büyük olduğu durumda bu yöntemin de yeterli olmayacağı tartışılmış ve TDA-tabanlı daha verimli bir yöntem önerilmiştir. Geleneksel yollarla ayrıştırılmasının mümkün olmadığı büyük katalog matrisi, rassallaştırılmış TDA ile ayrıştırılmıştır. Yer ve zamanda yüksek tasarruf sağlanırken başarımlar korunmuştur. Çoksesliliğin hakim olduğu kayıtlarda %76 gibi yüksek bir başarımlar sağlanmıştır. Çokseslilik derecesi bazında en az %65 başarımlar sağlanarak yöntemin tutarlı ve kararlı olduğu gösterilmiştir.

Büyük verilerin yaygınlaştığı çağımızda geleneksel yöntemlerin yetersiz kaldığı gösterilmiş ve çok büyük verilerde dahi uygulanabilir bir yöntem ile bu soruna çözüm getirilmiştir. Rassallaştırılmış matris ayrıştırmalarına dayalı yöntemler, devasa verikümeslerinin sıradanlaştığı çağımızda özellikle önemlidir. Yakın süreli araştırmalarımızın odağı bu yöndedir.

5. KAYNAKÇA

- [1] U. Şimşekli, "Bayesian Methods for Real-Time Pitch Tracking", MS Thesis, Boğaziçi Univ., İstanbul, 2010.
- [2] A. Klapuri ve M. Davy, "Signal Processing Methods for Music Transcription", Springer, 2006.
- [3] P. Smaragdis, "Polyphonic pitch tracking by example", 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 125–128. IEEE, 2011.
- [4] N. Halko, P. G. Martinsson, Y. Shkolnisky ve M. Tygert, "An Algorithm for the Principal Component Analysis of Large Data Sets", *SIAM Journal on Scientific Computing*, 33(5):2580, 2011.
- [5] G. Golub ve C. V. Loan, "Matrix computations", 3. basım. The Johns Hopkins University Press, 1996.
- [6] N. Halko, P. G. Martinsson ve J. A. Tropp, "Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions". *SIAM Review*, 53(2):217, 2011.
- [7] V. Emiya, R. Badeau ve B. David, "Multipitch Estimation of Piano Sounds Using a New Probabilistic Spectral Smoothness Principle", *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1643–1654, 2010.