Rassal Ormanlar, Değişken Seçimi ve Normal Yumuşatma Kullanarak Kahkaha ve Dolguların Tespiti

Heysem Kaya, Ali Mehdi Erçetin, Albert Ali Salah, Sadık Fikret Gürgen
Boğaziçi Üniversitesi, İstanbul
heysem@boun.edu.tr, mehdi.ercetin@boun.edu.tr, salah@boun.edu.tr, gurgen@boun.edu.tr

Özet: Bu çalışmada, konuşma içindeki kahkaha ve dolguların tespiti için çeşitli yapay öğrenme yöntemlerini irdelemekteyiz. Deneylerimizi Interspeech 2013 Paralinguistic Müsabakası, Sosyal İşaretler Alt-Müsabakası veri kümesi üzerinde gerçekleştirmekteyiz. Değişken seçimi için minimum Artıklık Maksimum İlgililik (minimum Redundancy Maximum Relevance - mRMR) süzgecini, sınıflama için Rassal Ormanları, ve son-işlemleme için Normal Yumuşatmayı kullanmaktayız. Bunun yanında, liner çekirdekli Destek Vektör Makinelerini (DVM) müsabaka makalesiyle uyumlu, bağımsız bir taban çizgisi elde etmek için kullanmaktayız. Sonuçlar Rassal Ormanların DVMlere hem müsabaka performans ölçütü Ağırlıklandırılmamış Ortalama Eğri Altında Alan (AOEAA) hem de eğitme verimliliği anlamındaki üstünlüğünü ortaya koymaktadır. Önerilen sistem ile bağımsız test kümesinde müsabaka taban çizgisini %5.1 (mutlak) düzeyde aşarak %88.4 lük AOEAA değerine erişmekteyiz.

Anahtar Sözcükler: Kahkaha/Dolgu Tespiti, Interspeech 2013 Sosyal İşaretler Alt Müsabakası, Sosyal İşaret İşleme, Rassal Orman sınıflandırıcı

Detection of Laughter and Fillers using Random Forests, Feature Selection and Gaussian Smoothing

Abstract: In this study, we investigate several machine learning methods for detection of laughter and fillers in speech. We experiment on Social Signals sub-challenge dataset from Interspeech 2013 Paralinguistic Challenge. We use minimum Redundancy Maximum Relevance (mRMR) filter for feature selection, Random Forests for classification, and Gaussian Smoothing for post-processing. We also use Support Vector Machines with linear kernel to provide an independent baseline in accordance with the challenge paper. The results indicate the superiority of Random Forests to SVMs in terms of the challenge performance measure, namely Unweighted Average Area Under Curve (UAAUC) as well as training efficiency. With the proposed system, we reach 88.4% UAAUC on independent test set, beating the challenge baseline by 5.1% (absolute).

Keywords: Laughter/Filler Detection, Interspeech 2013 Social Signals Sub-Challenge, Social Signal Processing, Human-Computer Interaction, Random Forest Classifier

1. Giriş

Bu çalışma, yazarların Interspeech 2013 uydu çalıştayı olan WASSS'ta yaptıkları bildiriye dayanmaktadır [1]. Çalışmadaki veri kümesi yine bu konferanstan edinilmiştir.

İşitsel-görsel tanıma teknolojisinin ulaştığı son noktanın olgunluğu, bilim insanlarının dikkatlerini duygusal ve sosyal işaretler gibi diğer konulara odaklamalarını sağlamıştır. En iyi sonuçlar farklı kiplerde, örn. konuşma ve görü, eğitilen modellerin kararlarının birleştirilmesi ile elde edilmektedir [2-8].

Interspeech 2013 Sosyal İşaretler Alt-Müsabakasında [9] dilsel olmayan olayların, diğer bir deyişle kahkaha ve dolguların, tespiti amaçlanmıştır. *Dolgular* konuşmacının yer tutmaya çalışmak için kullandığı "ah", "hımm", "eh" gibi seslendirmelerdir. Konuşma içinde kahkaha; eğlence, utanç ve coşku gibi hislerin güçlü ifadesidir. Bu nedenle konuşmacıya bağlı durumlarda dahi, farklı duygularla ilişkili seslendirmeler birbirinden ayrık akustik öznitelik değer kümeleri yansıtabilmektedir.

Müsabaka, kahkaha ve dolgunun sınıflandırılması ile bunların kalan "çöp" sınıfına karşı ayrımının çerçeve bazında yapılmasını dolayısıyla yer tespitini gerektirmektedir. Çerçeveler 10 milisaniye uzunluğunda ve örtüşmemektedir (detaylar için bkz. [9]). Meydan okuyan sorunların bir tanesi, konuşmanın büyük bölümü normal diyalog ve sessizlikten müteşekkil olduğu için çöp sınıfının hakimiyetidir. Sınıflandırılacak öğelerin sayısı da bir başka zorluk teşkil etmekdir ki burada ortaya çıkan bellek/hesaplama sorunu özenli öznitelik ve algoritma seçimiyle ele alınmalıdır.

İki sebepten dolayı müsabaka verileriyle birlikte sağlanan Technische Universitat Münchhen (TUM) referans öznitelik kümesini kullanmaktayız [9]. İlk olarak, sağlanan düşük düzeyli tanımlayıcılar - DDT (low level descriptor - LLD) örn. Mel Frekansı Kepstral Katsayıları (Mel Frequency Cepstral Coefficients - MFCC), F0, Harmoniklerin Gürültüye Oranı (Harmonics to Noise Ratio - HNR), ve türevlerin bu sınıflandırma amacına uygun olduğu bilinmektedir [3,5,10,11]. İkinci olarak, standart referans öznitelik kümesi tekrar edilebilirlik ve mukayese edilebilirliği artırmaktadır.

İnceleme altındaki sınıflandırıcı, Rassal Orman - RO (Random Forest - RF), biri öznitelik uzayında diğeri de örneklem uzayında olmak üzere iki rassallık kaynağı sunarak yapay öğrenmede oturmuş bir kavram olan *toplu öğrenme* (ensemble learning) gerçekleştirmektedir. Rassal Ormanlar her ağacın kendi rassal öznitelik kümesinde bilgi kazanımını esas alarak olarak uyguladığı sıralamayı kullandığı, büyük veri kümelerini bölüp-fethetmek için güçlü araçlardır. Bunun ötesinde, ağaç öğrenimi paralelleştirildiğinde öğrenme süreci herhangi bir başka global yapay öğrenme algorithmasından çok daha az zaman almaktadır.

Bu makalenin planı şu şekildedir: Bölüm 2'de kısaca ilgili literatürü sunmaktayız. Bölüm 3'te, veritabanı ve müsabaka referans özniteliklerini tanırmakta; ve sonra 4. bölümde, metodolojiyi vermekteyiz. Nihayet, bölüm 5 çalışmayı sonuçlandırmaktadır.

2. Literatür Taraması

Bu bölümde kahkaha tanıma problemi ile ilgili literatürü gözden geçirmekteyiz.

Ito ve diğerlerinin çalışmasında [3], ses öznitelikleri (MFCC ve ilk türevi) için Normal Bileşen Modelleri (NBM) , görsel (yüz) öznitelikleri için de bir liner diskriminant sınıflayıcı kullanılmıştır. Çok-kipli tümleştirme tek kipli kararların VE operatörüyle birleştirilmesi ile sağlanmıştır. Basitliğine rağmen birleştirme yönteminin doğruluğu artırdığı bulunmuştur.

Thruong and Leeuven Algısal Liner Tahmin (ALT) ve vezin özniteliklerini çerçeve düzeyinde; temel frekans, seslendirme ve modülasyon spektrumu özniteliklerini söyleyiş düzeyinde özütlemiştir [4]. NBM ve Destek Vektör Makinaları (DVM) esas sınıflayıcılar olarak kullanılmış, akabinde kararlar TOPLAM kuralı, Çok Katmanlı Algılamaç (ÇKA) sinir ağı veya DVM ile tümleştirilmiştir. Sonuçlar farklı sınıflayıcıların karar tümleştirilmesinin (örn. DVM ve NBM) aynı algoritma farklı öznitelikli sınıflayıcıların kararlar tümleştirmesine göre daha iyi olduğunu göstermiştir.

Knox and Mirghafori çerçeve düzeyinde kahkaha tespiti için tek kipli bir çalışma yapmıştır [5]. Çalışmada MFCC ve vezin öznitelikleri (temel frekans ve enerji) ilk ve ikinci nümerik türevle-

riyle birlikte kullanılmıştır. Her bir çerçeve için verilecek karar, ilgili çerçeve ve komşularıyla beraber 75-çerçeveden elde edilen DDTler bir sinir ağında eğitilmiştir. Her DDT türü türev öznitelikleriyle birlikte ayrı bir yapay ağında eğitilmiştir. Sonrasında DDT-tabanlı sinir ağlarının sınıf sonsal olasılıkları bir başka sinir ağına istiflenmiştir. Çalışma, ilgili tarihte, ICSI Toplantı Veritabanı – BMR altkümesi üzerinde en iyi sonuçlara ulaşmıştır. Kayda değer bir bulgu, tekbaşına MFCC türev özniteliklerinin ham MFCC özniteliklerinden daha iyi sonuçlar vermesidir.

Aynı problem üzerindeki iki ayrı çalışmalarında, Petridis ve Pantic ses ve görüntü özniteliklerinin karar tümleştirmesini basit TOPLAM kuralı ve sinir ağına istifleme yaparak irdelemiştir [6,7]. Görsel işaretlerin ALT, temel frekans ve enerji temelli ses özniteliklerden daha üstün olduğu gösterilmiştir. Beklendiği üzere, çok-kipli tümleştirme tek-kipli modellerden daha iyi sonuç vermiştir. Daha ötesinde, en iyi sonuçlar tek-kipli kararların sinir ağına istiflenmesiyle elde edilmiştir.

2009 yılında yaptıkları literatür çalışmasında, Vinciarelli ve diğerleri alanın yeniliğinin altını çizerken sosyoloji, psikoloji, bilgisayar mühendisliği alanlarının ve bir o kadar da yapay öğrenme ve işaret işleme bilimleri gibi dalların ortak çalışmasını gerektirdiğini ifade etmiştir [8]. Bununla birlikte çalışma sosyal işaretlerin doğal algısının çok-kipliliğine ve bunun otomatik tanımaya yansıması gerektiğine vurgu yapmıştır. Çok-kipli birleştirme, ve hatta tek-kipli değişik sınıflayıcıların tümleştirilmesinin gürbüz ve doğru sonuçlar verdiği bilinmektedir.

Bildiğimiz üzere, içsel olarak sınıflayıcı birleştirme özelliği bulunan Rassal Ormanlar kahkaha tespiti problemine uygulanmamıştır.

3. Sosyal İşaretler Alt-Müsabakası: Veritabanı ve Öznitelikler

INTERSPEECH 2013 Sosyal İşaretler Alt-Müsabakası "SSPNet Selendirme Veritabanı" ("SSPNet Vocalization Corpus" - SVC). Okuyucuyu veritabanı ve çıkarılan öznitelikler hakkında bilgi sahibi yapmak için üzeirnden geçmeteyiz. Diğer tüm detaylar Schuller ve diğerlerinin makalesinde [9] bulunabilir. Tablo 1 SVC veritabanının temel istatistiklerini vermektedir: Toplamda, her biri 11 saniye uzunuğunda 2763 klip, 1.2 K kahkaha olayı ve 3.0 K dolgu olayı vardır. Veritabanı 120 kişi arasındaki 60 telefon görüşmesinden yapılmıştır. 120 kişinin 63'ü kadındır. Müsabaka için kişi bağımsızlığı sağlama amacıyla [9], 70 kişiye karşılık gelen ilk 35 görüşmeler *öğrenme*, 36 ila 45 numaralı görüşmeler *geliştirme* ve kalanlar (46-60) *test* kümesi olarak ayarlanmıştır.

Özellik	İstatistik
Klip Sayısı	2763
Klip Süresi	11 sn.
Telefon Görüşme Sayısı	60
Kişi Sayısı	120
Erkek Katılımcı Sayısı	57
Kadın Katılımcı Sayısı	63
Dolgu Olay Sayısı	3.0 k
Kahkaha Olay Sayısı	1.2 k

Tablo 1. SVC Veritabanı Özet İstatistikleri

Problemde amaç 10 milisaniyelik örtüşmeyen çerçeveleri dolgu, kahkaha ve çöp sınıflarından birine atamaktır. Veritabanında 10 ms lik çerçevelerin toplam sayısı 3 027 949'dur, bu da yüksek örneklem zorluğu yaratmaktadır. Öğrenme kümseinde 59 294 kahkaha çerçevesi , 85 034 dolgu çerçevesi ve 1 591 442 çöp sınıfı çerçevesi bulunmaktadır. Alt-örneklemleme önlemi olarak müsabaka makalesi öğrenme kümesindeki her 20 'çöp' çerçeveden birini almayı önermektedir, bu da öğrenme çerçevelerini 79 572'e indirmektedir [9].

Yüksek örnek sayısından kaynaklanan bellek kısıtlarını dikkate alarak duygusal anlamda yüksek potansiyel taşıyan az sayıda bir öznitelik kümesi çıkarılmıştır [9]. Çerçeve düzeyinde çıkarılan bu DDTler MFCC 1-12, logaritmik enerji, seslendirme olasılığı , Harmoniklerin Gürültüye Oranı (HGO), F0, işaretin sıfır

geçme oranı ile bunların birinci türevlerinden oluşmaktadır. MFCC and logaritmik enerji için ikinci derece türev katsayıları da çıkarılmıştır. Bu çerçeve düzeyindeki DDTler 8 komşu çerçevenin (4 önce ve 4 sonra) aritmetik ortalama ve standart sapma istatistiği ile bir araya getirilip toplamda 141 öznitelik elde edilmiştir [9].

4. Metodoloji

Sınıflandırma için Rassal Ormanları kullandık [12]. Rassal Ormanlar karar ağaçlarının birleştirilmesi ile elde edilir (karar ağaçlarının detaylarının detayları için bkz. [13] böl. 9). Her karar ağacı veriden rasgetgele seçilen bir örneklem kümesi (yeniden yerleştirilerek örneklenen) ve yine rastgele seçilen bir öznitelik kümesi ile eğitilir. Yeniden yerleştirme ile rassal örnekleme her ağaç için ortalamada örnek sayısının 1/3 ü kadarını 'çanta dışında' bırakır. Bu çanta dışı örnekler eğitim zamanında 'genelleştirme hatasının tarafsız tahmini' için kullanılır [12]. Orman bireysel ağaçların modu (en yüksek frekansa sahip değer) olan sınıfı çıktılar. Rassal Ormanların büyük veritabanlarında verimli çalışmaları ve mevcut algoritmalara üstünlükleri gösterildiği için eldeki probleme uygun olduğunu düşünmekteyiz [12].

Boyutsallığın azaltımında minimum Artıklık Maksimum İlgililik (minimum Redundancy Maximum Relevance - mRMR) değişken seçim yöntemini kullandık [13]. mRMR problemi temsil eden en düşük kümesi bulma gayretiyle, her defasında hedef değişken ile en çok ilgili olan ancak seçili değişkenlerle en düşük ortak bilgiye (artıklık) sahip değişkenin seçimini önerir. $\{S_i\}$ hazlihazırda seçili değişken kümesini ve *y* hedef değişkeni, *I(a,b) ise a* ve *b* vektörleri arasındaki ortak bilgiyi göstersin; mRMR, *I(f,y) ile* $\sum I(f,S_i)$ farkını yahut oranını en yüksek yapan f değişkenini seçer.

Referansımız müsabaka makalesinde de kullanılan liner DVM sınıflayıcıdır [2] (DVM detayları için [13] böl. 13'e başvurabilirsiniz). Bağımsız taban çizgisi elde etmek için [9]'deki deneyleri tekrar ediyoruz.

4.1 Deneysel Sonuçlar

DVM ve RO lar için Weka [14] uygulamasını kullandık. mRMR değişken seçimi için Peng ve diğerlerinin orijinal uygulamasından faydalandık [15]. Rassallık içeren bütün deneylerimizde çekirdek (seed) değerini tekrar edilebilirliği kolaylaştırmak için varsayılan değer olan 1'e atadık.

Müsabaka makalesinde [9] önerildiği şekilde, sosyal işaretler alt-müsabakası için kahkaha ve dolgu sınıflarının Eğri Altındaki Alanı (EAA) ve bunların ağırlıksız ortalaması (AOEAA) kullanılan performans ölçüsüdür.

4.1.1 Taban Çizgisinin Yeniden Elde Edilmesi

Müsabaka makalesinde rapor edilen geliştirme kümesi EAA performansı (C=0.1, DVM liner çekirdek) sırasıyla kahkaha ve dolgu için %86.2 ve %89.0'dır. Müsabaka makalesinde kullanıldığı gibi alt-örneklemleme yapılmış veri kümesi üzerinde deneyler yaptık. Ancak, aynı kurgu ile tekrar edilen deneyler aynı EAA değerlerini vermemiştir. (bkz. Tablo 1). Bu durumu rapor edilen performansın alt-örneklemleme ile elde edilen kümeden değil tüm veriden çıkarılan sonuç olması olasılığına atfettik. Kendi çalışmalarımıza referans teşkil etmesi için bütün değişken kümesi ile DVM linear kernel ve ceza parametresi C = {0.1, 1, 10} ile testler yaptık. Tablo 2'de görüleceği üzere, her 3 DVM ceza değeri aynı AOEAA sonucunu verdi.

C / %	Kahkaha	Dolgu	AOEAA
0.1	81.3	83.6	82.5
1	81.2	83.7	82.5
10	81.2	83.7	82.5

Tablo 2. Geliştirme Kümesi DVM Taban Çizgisi

Bunun ötesinde mRMR filtresinin etkisini en yüksek dereceli 50, 70 ve 90 değişken ile değerlendirdik. mRMR değişken seçimini uygulamak için Peng ve diğerlerinin [15] önerdiği şekilde bir ayrıklaştırma yapıldı. Bu amaçla, her bir değişken için ayrı ayrı z-skor hesaplanıp bu skorun onda biri en yakın tamsayıya yuvarlandı (diğer bir ifadeyle her bir standart sapma 10 dilime ayrıldı).

DVM üst parametresi C ye bağlı bir fark görülmediği için C=0.1 ile elde eidlen sonuçları veriyoruz. Bulgular seçilen değişken sayısı azaldıkça DVM liner çekirdek kullanarak yapılan tespit oranının düştüğünü göstermektedir (bkz. Tablo 3). Bütün diğer parametreler sabit tutulurken 141 orijinal özniteliğin 50'si kullanıldığında AOEAA'daki düşüş %2 (mutlak) düzeydedir.

# mRMR	Kahkaha	Dolgu	AOEAA
90	80.7	83.3	82.0
70	79.9	83.0	81.5
50	78.8	82.4	80.6

Tablo 3. Geliştirme Kümesi üzerinde Liner DVM ve mRMR Süzgeci Performansı *(%EAA)*

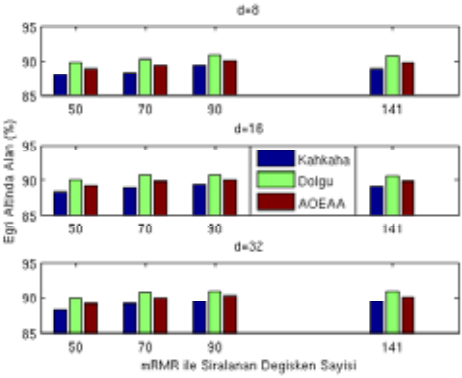
DVM ile yapılan sonraki çalışmalarda değişken kümesinin bölümlenmesi ve ayrı eğitilen modellerin karar birleştirmesi sonuçları geliştirmemiştir Tablo 2'de verilen taban çizgisinin üzerine çıkarmamıştır. Benzer şekilde DVM'nin RBF çekirdeği ile eğitimi daha uzun zaman almış ancak yine istenen performansı verememiştir. Deneysel gözlemlerimize göre C 10 katına çıkarıldığında eğitim süresi ortalama 10 kat artmaktadır. C değeri 10 alındığında, 50 değişken ile eğitim $100x10^3$ saniyeden fazla sürmüştür.

4.1.2 Rassal Ormanlar ile Deneyler

Rassal Ormanların diğer algoritmalardan daha iyi performans göstermeleri ve büyük veritabanlarında verimli çalışmaları [12] sebebiyle bu probleme Destek Vektör Makinalarından daha uygun oldukları öngörülmüştür. Rassal ormanların üst-parametreleri ağaç sayısı (*T*) ve her ağacın eğitiminde rassal olarak seçilecek değişken sayısından (*d*) ibarettir.

Bununla birlikte DVM deneylerinde olduğu gibi mRMR ile sıralanmış değişkenlerin (*D*) etkisini de inceledik. Üst parametler için test edilen değer kümeleri d = {8, 16, 32}, T= {10, 20, 30} ve D = {50, 70, 90, Tümü} şeklindedir. Dikkat edilirse $d \leq D$ olması gerektiğinden, *d* için daha yüksek değerler test edilmemiştir.

ROlardan elde edilen AOEAA sonuçları DVMlerden daha iyi olduğu görülmüştür. En iyi DVM/ OEAA taban çizgisi performansı %82.5 iken, en kötü (*d*=8, *D*=50, *T*=10 ile eğitilen) RO performansı %86.3 lük bir AOEAA performansı göstermiştir. Figür 1'de 20 ağaçla farklı yerel değişken sayısı (d) ve global değişken sayısı (D) ile eğitilen ROların performansları gösterilmiştir. Figürden de görüleceği üzere ROların performansı genel olarak DVM'lerden daha iyidir. Bunun da ötesinde d=16 iken, global değişken sayısının mRMR ile yarıya indirilmesi (D=70) AOEAA performansını olumsuz etkilememektedir.



Figür 1. *Geliştirme Kümesinde T=20, değişen d ve D Parametreleriyle RO Performansı*

Tablo 4 değişik d, D ve T üst parametreleriyle yapılan testlerdeki AOEAA performansını özetlemektedir. AOEAA her üç üst-parametreye (*d, D* ve *T)* bağlı olarak artmaktadır. *T*=10'dan *T*=20'ye ortalama AOEAA performans artışı (1.5%), *T*=30 ile *T*=20 arasındaki farktan (0.3%) daha büyüktür. ROların aşırı öğrenme yapmadığı bilinse de [12], ağaç sayısı artışına karşı performans asimtotik davranmaktadır. Bununla birlikte, D ye bağlı ortalama performans düşümü (141 den en iyi mRMR 50'ye) DVM testlerinde %2 iken ROlarda %1 olarak gerçekleşmiştir. Daha ötesinde, mRMR ile sıralanan ilk 90 değişken kullanıldığında (**koyu** ile gösterildi) tüm değişkenlerin kullanıldığı modellerden nispeten daha iyi sonuçlar elde etmek mümkün olmuştur.

d	D	10 Ağaç	20 Ağaç	30 Ağaç
8	Tüm	87.4	89.8	89.7
	90	**88.0**	**90.1**	**90.1**
	70	87.9	89.3	89.8
	50	87.5	88.9	89.4
16	Tüm	88.2	89.8	90.4
	90	**88.6**	**90.0**	**90.5**
	70	88.5	89.8	90.2
	50	87.8	89.2	89.6
32	Tüm	88.8	90.2	90.4
	90	**89.0**	**90.3**	**90.8**
	70	88.7	90.0	90.4
	50	87.9	89.1	89.6
Ort		88.2	89.7	90.0

Tablo 4. *Geliştirme Kümesinde Değişen T, d ve D Parametreleriyle RO Performansı (%AOEAA)*

Çalışmaları daha ileri götürerek sonsal olasılık dağılımlarına çerçeve komşuluk bilgilerini kullanarak Normal yumuşatmanın etkisini irdeledik. Buradaki fikir milisaniyeler içinde sınıflar arasında keskin geçişlerin (örn. kahkahadan konuşmaya sonra geri kahkahaya) olası olmadığı idi. Bu amaçla sınıf sonsal dağılımları her çerçevenin 2K komşusu ile (K öcesi ve K sonrası) yumuşatıldı. K için 1-10 arası değerler denenmiştir. *i* indisli çerçeveyi komşu j indisli, $i-K \le j \le i+K$, çerçeve ile yumuşatmak için kullanılan Normal ağırlık fonksiyonu şu şekilde verilmiştir:

$$w_{i,j} = (2 * pi * B)^{-(1/2)} * \exp\left(-(i-j)/(2*B)\right) \quad (1)$$

burada |i-j| çerçeve zaman indislerinin farkının mutlak değeridir. Sadelik amacıyla testlerimizde B değeri 1 olarak kullanılmıştır. K ya göre performans artışını göreceli değerlendirmek için gelşitirme kümesindeki AOEAA yerine hesaplanması daha kolay olan tahmin başarısını ölçtük. K=8'dan K=9'a tahmin başarısı %0.05 , K değeri 8 olarak belirlenmiştir. Bu değerle yumuşatılan sonsal olasılıkların EAA ölçüsü Kahkaha ve Dolgu sınıfları için sırasıyla 92.2 ve 92.4 olmuştur. Bu, geliştirme kümesinde %1.5 (mutlak) düzeyde bir artışa karşılık gelmektedir.

Geliştirme kümesindeki en iyi model kurgusu ile test kümesi performansını ölçmek için eğitim ve geliştirme kümelerini birleştirerek her biri global (mRMR ile seçilen) 90 değişken içinden rasgele 32 değişken seçerek büyütülen 30 ağaçlı bir orman eğittik. Sonrasına K=8 ile (16 komşu çerçeve kullanılarak) sonsal olasılıkları yumuşattık. Test kümesinde kahkaha ve dolgu için sırasıyla %89.6 ve %87.3 EAA skorları elde ettik. Sonuç AOEAA (%88.4) müsabaka test kümesi performansını (%83.3) %5.1 mutlak oranda geçmiştir.

5. Sonuç ve Öneriler

Bu çalışmada, kahkaha ve dolgu tespiti için Rassal Ormanların kullanımını önerdik. Sonuçlarımız hem tahmin başarısı hem de verimlilik (özellikle ağaç öğrenimi paralelleştirilirse) anlamında ROların üstülüğünü ortaya koymuştur. Öğrenme yapılarından dolayı ROlar genelleştirme hatasının tarafsız kestirimini öğrenme anında sağlama özelliğine sahiptir. Bu müsabakadaki ana zorluk çerçeve bazında kahkaha-dolgu tahmini gerekliliğidir. Bu durum kullanılacak değişken sayısını büyük ölçüde kısıtlamaktadır.

Bu makalede, müsabaka verisiyle ile birlikte sunulan standart referans öznitelik kümesini değerlendirdik. Müsabakada verilene benzer bir taban çizgisi elde etmek için, liner çekirdekli DVM kullandık. Son zamanların gözde değişken seçim yöntemi mRMR ile değişken sayısının azaltımının imkan ve etkilerini inceledik. DVMlerin nispeten zayıf performansı ve uzun eğitim süresi bizi problemide daha verimli çalışacağını düşündüğümüz ROları irdelemeye teşvik etti. Beklediğimiz üzere ROlar ile daha iyi sonuçlar elde ettik. Geliştirme kümesinde, ilk 90 mRMR seçimli değişken ile 141 değişkenden nispeten daha iyi sonuçlar elde ettik. Dahası, Normal yumuşatma ile sonişlemlemenin performansa katkıda bulunduğunu gözlemledik. 16 komşu çerçeveyi Normal yumuşatmada kullanarak AOEAA ölçüsü %1.5 düzeyinde iyileşti. Geliştirme kümesinde, toplamda müsabaka makalesindeki kurguda (liner

çekirdekli DVM, C=0.1). verilen taban çizgisini tekrar ettiğimiz değerden %9.8 daha yükseğe çıkabildik. Son olarak en iyi üst parametre kurgusuyla eğitim ve geliştirme kümelerini birlikte eğitip Normal yumuşatma uyguladık. Bu modelle test kümesi AOEAA taban çizgisi üzerinden %5.1'lik bir artışla %88.4 düzeyinde performansa ulaşılmıştır.

İlerki çalışmalarda ses işarettinden çıkarılan başka özniteliklerin kullanımı ve farklı kahkaha/dulgu tiplerini modellemek için bileşen modellerinden faydalanmayı planlıyoruz. Normal yumutşatmada kullanılan B parametresinin başarıya etkisini de incelemeyi planlamaktayız.

6. Kaynaklar

[1] Kaya, H., Erçetin A. M., Salah, A. A., Gürgen, F., "Random Forests for Laughter Detection", Proceedings of Workshop on Affective Social Speech Signals - in conjunction with the INTERSPEECH 2013, Grenoble, France, 2013

[2] Dupont, S., & Luettin, J., "Audio-visual speech modeling for continuous speech recognition", IEEE Transactions on Multimedia, 2(3), 141-151, 2000

[3] Ito, A., Xinyue, W., Suzuki, M., and Makino, S., "Smile and Laughter Recognition using Speech Processing and Face Recognition from Conversation Video", In Proceedings of the 2005 International Conference on Cyberworlds, Washington, DC, USA, 437-444, 2005

[4] Truong, K. P., and Van Leeuwen, D. A., "Automatic discrimination between laughter and speech", Speech Communication, 49(2), 144-158, 2007

[5] Knox, M., and Mirghafori, N., "Automatic laughter detection using neural networks", Proc. Interspeech 2007, 2973-2976, 2007

[6] Petridis, S., and Pantic, M., "Audiovisual discrimination between laughter and speech", In Acoustics, Speech and Signal Processing ICASSP 2008. IEEE International Conference on, 5117-5120, 2008

[7] Petridis, S., and Pantic, M., "Fusion of audio and visual cues for laughter detection", In Proc. 2008 International conference on Content-based image and video retrieval (CIVR '08), ACM, New York, NY, USA, 329-338, 2008

[8] Vinciarelli, A., Pantic, M., Bourlard, H., "Social signal processing: Survey of an emerging domain", Image and Vision Computing 27.12 1743-1759, 2009

[9] Schuller, B., Steidl S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., Chetouani, M., Weninger, F., Eyben, F., Marchi, E., Salamin, H., Polychroniou, A., Valente F. and Kim S., "The Interspeech 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism", Proc. Interspeech 2013, ISCA, Lyon, France, 2013

[10] Schuller B., In Salah, A. A. and Gevers, T. (eds) "Computer Analysis of Human Behavior" chap. Voice and Speech Analysis in Search of States and Traits, 227–253, Springer, 2011.

[11] El Ayadi, M., Kamel , M. S., and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases", Pattern Recognition, 44(3), 572–587, Mar. 2011

[12] L. Breiman, "Random Forests", University of California, Berkeley,USA, 2001

[13] Alpaydın E., "Introduction to Machine Learning". Massachusetts, USA: The MIT Press, 2010

[14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, "The WEKA Data Mining Software: An Update; SIGKDD Explorations", 11(1), 2009.

[15] Peng, H., Long, F., and Ding, C., "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy", IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8), 1226-1238, 2005