



FARKLI VERİ SETLERİ ARASINDA DUYGU TANIMA ÇALIŞMASI (*EMOTION RECOGNITION STUDY BETWEEN DIFFERENT DATA SETS*)

Cevahir PARLAK¹, Banu DİRİ²

ÖZET/ABSTRACT

İnsanlar arasındaki en önemli iletişim aracı konuşmadır. Konuşma ile insanlar birbirlerine sadece düşüncelerini değil duygularını da aktarabilirler. Konuşma ile karşımızdaki kişinin düşüncesini, duygusunu, cinsiyetini ve yaşını da tahmin edebilir. Bu çalışmada EmoSTAR adlı yeni bir duygu veri seti sunulmuş ve Berlin Duygu Veri seti ile çapraz testler yapılmıştır. Çapraz testlerde, setlerden biri eğitim diğeri test seti olarak kullanılmıştır. Ayrıca, çalışmada özellik seçicilerin performansı da incelenmiştir. Özellik çıkarımı için openSMILE Emobase ve Emo_large konfigürasyonlarında MFCC sayısı 12'den 24'e çıkartılarak ve Harmonik Gürültü Oranı özellikleri eklenerek gerçekleştirilmiştir. Özellik seçme ve sınıflandırma ise Weka aracıyla yapılmıştır. EmoSTAR halen daha fazla duygu türü ve örnek için geliştirilme aşamasındadır.

Speaking is the most important communication tool between people. By speaking people can transfer not only their thoughts but also their feelings to each other, too. When speaking we can estimate the conception, feeling, gender and age of the person we are talking to. In this study, a new sense of a data set called EmoSTAR is presented and cross tests with Berlin emotions data set is made. In the cross-test, one of the data set is used as training set while the other data set is used as test set. Additionally, in the study the performance of feature selector has been also examined. For feature extraction MFCC number is increased from 12 to 24 in openSMILE Emobase and Emo_large configurations and also developed by adding the Harmonic-to-Noise-Ratio features. The feature selection and classification is made by the Weka tool. EmoSTAR is currently under development for more emotion and sample type.

ANAHTAR KELİMELER/KEYWORDS

Çok setli duygu analizi, Duygu madenciliği, Prozodi, EmoDB, EmoSTAR
Cross-corpus emotion analysis, Emotion mining, Prosody, EmoDB, EmoSTAR

¹ Beykent Üniversitesi, Bilgisayar Programcılığı, İstanbul, cevahirparlak@yahoo.com

² Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği, İstanbul, banu@ce.yildiz.edu.tr

1. GİRİŞ

Duygu insanları diğer canlılar ile makinelerden ayıran ve insanlar arasındaki iletişimde çok önemli rol oynayan bir faktördür. Son yıllarda insan makine etkileşiminde duygu tanıma çalışmaları oldukça ilerleme kaydetmiştir. Makinelerin ürettiği mekanik sesli cevaplar insanlar üzerinde itici bir etki oluşturmaktadır. Bu nedenle insan-makine etkileşiminde makinelere duygu yüklü konuşmalar yapabilme ve karşısındakinin duygusunu anlayabilme özelliği kazandırılmaya çalışılmaktadır. Böylece insan-makine etkileşimi daha çekici olabilmektedir. Duygu tanıma çalışmalarının temel öğelerinden birisi de hiç kuşkusuz duygu veritabanlarıdır. Ne yazık ki bu konuda üstünde çalışılacak veri setlerini hazırlamak oldukça zahmetli ve zorlu bir çalışma gerektirmektedir. Pek çok duygu türü bulunmakta ve bu duygu türlerini gerçek ortamlarda elde edebilmek çok ciddi problemleri beraberinde getirmektedir. Bu konudaki ilk çalışmalarda daha az duygu türleri üstünde veya konuşmanın sadece olumlu veya olumsuz olması ile ilgilenilmekte iken, gelişen teknolojiyle birlikte daha fazla duygu türü üstünde çalışılmaya başlanmıştır.

Duygu tanımanın uygulama alanları arasında çağrı merkezi uygulamaları, bilgisayar destekli öğrenme sistemleri, yalan makineleri, sesli e-posta sistemleri ve oyunlar sayılabilir. Son yıllarda otomobillerde de sürücünün performansı ile duygu hali arasındaki ilişkiyi gözlemlemek amacıyla duygu tanımadan yararlanılmaktadır (Ramakrishnan, 2012). Duygusal konuşma sentezi alanında ki çalışmalar da gün geçtikçe artmaktadır (Black vd., 2011; Iida vd., 2003).

2. İLGİLİ ÇALIŞMALAR

Duygu tanıma ile ilgili ilk araştırmalara 80'li yılların ortalarından itibaren başlanmış ve bilgisayar teknolojisindeki gelişmelere paralel olarak 90'lı yılların ortalarından itibaren hız kazanmıştır. Bu çalışmalar önceleri sadece konuşmanın olumlu veya olumsuz oluşu ile ilgilenilmekte iken, sonraları teknolojinin gelişmesiyle beraber pek çok duygu türünü ve stres derecesini de kapsama alanı içine almıştır. Bu çalışmaların temel elemanları duygu veritabanlarıdır. Duygu tanıma çalışmalarında kullanılan veritabanlarındaki duygu türü ve bu duyguları sınıflandırmak için kullanılan özelliklerin sayısı araştırmalara göre birbirlerinden çok farklılık göstermektedir. Bazı araştırmalarda hedeflenen duygu sayısı az olduğundan sayılı birkaç özellik kullanılırken, bazı araştırmalarda ise binlerce özelliğin kullanılması gerekmektedir. Bu tür çalışmalarda önceleri klasik özellikler olarak bilinen prozodik özellikler kullanılır iken, bugün artık prozodi ve spektrum özelliklerinin yanı sıra bunlardan çeşitli istatistiksel fonksiyonlarla türetilen binlerce özellik de kullanılmaktadır. Bu çalışmalarda duygu sınıflandırmanın yanı sıra stres sınıflandırma ve duygu yüklü konuşma sentezi de yapılmaktadır. Son zamanlarda sesin akustik özellikleriyle beraber kelime tanıma ve yüz ifadesi tanıma da kullanılmaya başlanmıştır. Sınıflandırmadaki başarı oranları da veritabanına ve duygu sayısına göre % 30 ile % 90 arasında farklılıklar göstermektedir.

Duygu tanıma çalışmaları genellikle tek bir veritabanı üstünde gerçekleştirilmektedir ve başarı oranlarında limitlere yaklaşmıştır. Bu nedenle bu çalışmada iki farklı veritabanı ile çalışılacak ve bu veritabanlarından biri eğitim seti olarak kullanılırken, diğeri test seti olarak kullanılmak suretiyle çapraz testler gerçekleştirilecektir. Bu tür çalışmalar literatürde nispeten daha yeni ve az olduğu gibi üstesinden gelinmesi gereken fazla problemi içerisinde barındırmaktadır. Çapraz testli çok setli çalışmalarda ise başarı oranları tek setli çalışmalara göre oldukça düşük çıkabilmektedir.

2.1 Çok Setli Mevcut Çalışmalar

Oflazoğlu ve Yıldırım, TURES (Türkçe Duygusal Konuşma Veritabanı) ve VAM (Vera am Mittag) veri setlerinde setler arası bir çalışma gerçekleştirmiş olup, 3 boyutlu duygu uzayında TURES'in eğitim seti olarak kullanılmasıyla 4 duygu (kızgın, nötr, mutlu, üzgün) üstünde pozitiflik, aktivasyon ve baskınlık için sırasıyla % 0,28, % 0,74 ve % 0,71 korelasyon katsayıları elde etmişlerdir (Oflazoğlu ve Yıldırım, 2013). Bu çalışmada openSMILE ile çıkarılan 1532 özellik TURES ve VAM veritabanlarında kullanılmıştır. VAM ve TURES veritabanları ayrı ayrı test ve eğitim verisi olarak kullanılmış ve çapraz uygulama yapılmıştır. Sadece TURES veritabanı üstündeki tek setli çalışmada ise, 4 duygu (kızgın, nötr, mutlu, üzgün) üstünde Weka LIBSVM sınıflandırma ile % 57,5 başarı oranına ulaşılmıştır.

Zhang vd., ABC, AVIC, DES, eINTERFACE, SAL ve VAM setlerinde yaptıkları çapraz testlerde iki boyutlu duygu uzayında, aktivitede % 62,6 ve pozitiflikte ise % 55,6 başarı oranı elde etmişlerdir (Zhang vd., 2011).

Schuller vd., EmoDB'nin test seti, AVIC, DES, eINTERFACE, SmartKOM ve SUSAS'ın eğitim seti olarak kullanıldığı çalışmalarında % 35 ile % 45 arasında başarı oranları elde etmişlerdir (Schuller vd., 2010).

ABC, AVIC, DES, EmoDB, eINTERFACE, SAL, SUSAS ve VAM olmak üzere 8 veri setinin kullanıldığı bir diğer çalışmada VAM üstünde eğitim, diğer 7 veri seti üstünde de test yapılarak iki boyutlu duygu uzayında aktivitede % 67,7 başarı oranı yakalanmıştır (Schüller vd., 2011). DES üstünde eğitim, diğer 7 veri seti üstünde test yapılarak da pozitiflik boyutunda % 54,8 başarı oranı yakalanmıştır.

Eyben vd., SVM (SMO) ile SmartKOM, Aibo, SAL ve VAM setleri üstündeki çalışmada setlerden üçünü eğitim, diğerini de test seti olarak kullanarak % 53,4 başarı oranı elde etmişlerdir (Eyben vd., 2010).

Schuller vd., çalışmasında, EmoDB, ABC, AVIC, DES, SUSAS, Enterface, SAL, SmartKom ve VAM veritabanlarında 6552 özellikten oluşan bir konfigürasyon denenmiş ve EmoDB'de HMM/GMM tabanlı sınıflandırıcı ile % 77,1 ile en yüksek, SAL veritabanında ise % 32,7 ile en düşük başarı oranı gözlemlenmiştir (Schuller vd., 2009). Aynı çalışma, SVM ile EmoDB'de % 85,6 ile en yüksek, SAL'da % 30,6 ile en düşük oranı elde etmiştir.

2.2 Tek Setli Mevcut Çalışmalar

He, SUAS veritabanında 3 düzeyli stres tanıma üzerinde çalışmıştır. TEO tabanlı özelliklerle stres tespitinde % 92, ORI veritabanında TEO tabanlı özelliklerle 5 duygu üstünde % 87 başarı oranı elde etmiştir (He, 2010).

Bhargava ve Polzehl ise, MFCC ve prozodik özelliklere ek olarak konuşma ritmi ve ses yüksekliği özelliklerini de kullanılarak, 487 özellikle SVM ile Berlin EmoDB'de 7 duygu üstünde % 80,27 başarı oranı elde etmiştir (Bhargava ve Polzehl, 2012). Duygular ikili gruplar halinde sınıflandırılmaya tabi tutulmuştur.

Batliner vd., doğal bir veritabanı olan FAU Aibo üzerinde, 3713 akustik ve 531 lingistik özellikle 4 duygu üstünde yaptığı çalışmasında sadece akustik özelliklerle % 63,4 ve sadece lingistik özelliklerle de % 62,6 başarı oranı elde etmiştir (Batliner vd., 2009). Akustik ve lingistik özelliklerin beraber kullanılmasıyla da % 65,35 başarı elde etmiştir.

Shahzadi vd.'deki çalışmasında 204 spektral örüntü, 780 Harmonik Enerji, 241 prozodik ve 960 spektral olmak üzere toplam 2185 özellik kullanmıştır (Shahzadi vd., 2013). Spektral ve Harmonik Enerji özelliklerinin örüntülerini kullanarak EmoDB'de 7 duygu için % 86,9 başarı oranı elde etmiştir. Kadın ve erkek sesleri üstündeki çalışmalar ise ayrı ayrı değerlendirmeye tabi tutulmuştur.

Chavhan vd.'de 12 Mel Filtre bankasının 1. ve 2. farkları alınarak yeni özellikler türetilmiştir ve LibSVM ile kızgın, nötr, mutlu, üzgün ve korku olmak üzere 5 duygu üstünde % 93,75'lik başarı oranı elde etmişlerdir (Chavhan vd., 2010). Cinsiyet tabanlı sınıflandırmada ise erkeklerde % 94,73, kadınlarda ise % 100 başarı elde edilmiştir.

Wu vd. ise, zaman-frekans özelliklerini kullanarak EmoDB üstünde 7 duygu sınıfında modülasyon frekans özelliklerini kullanarak % 85,6 başarı elde etmiştir (Wu vd., 2010).

Schuller vd., FAU Aibo veritabanı üstünde farklı özellik setlerinin başarısını karşılaştırmıştır (Schuller vd., 2007). SVM ve Rassal Orman sınıflandırıcıları tüm özellikler ve indirgenmiş 150 özellik ile % 50 ile % 64 arasında başarı oranı elde etmişlerdir. Özellik gruplarındaki özellik sayısı 26 ile 1718 arasında değişmektedir.

Mairesse vd., 84 katılımcı ile bir çalışma gerçekleştirmiş ve katılımcılara gittikleri restoranlarla ilgili düşüncelerini cep telefonlarıyla yaptıkları konuşmalarda belirtmelerini istemiş ve düşünceleri olumlu ve olumsuz olarak sınıflandırmıştır (Mairesse vd. 2012). Akustik özelliklerle beraber konuşma tanıma özelliğinden de yararlanmış ve ayrı ayrı performansları ölçülmüştür. Temel frekansa bağlı özelliklerin olumlu/olumsuz sınıflandırılmasında % 72,6 ile en yüksek başarı oranını almıştır. Bu çalışmada konuşma tanımadaki hataların, başarı oranını fazla etkilemediği görülmüştür.

Wu vd., çalışmasında VAM veritabanı üstünde 46 akustik özellik kullanmıştır (Wu vd., 2011). Bunların 26'sı MFCC, 6'sı enerji, 9'u temel frekans ve 5'i süre özelliğidir. Bu özelliklerin duygu tanımda önemleri araştırılmış ve MFCC özelliklerinin en yüksek önceliğe sahip olduğu temel frekans ve enerjinin de onu takip ettiği görülmüştür. Süre ile ilgili özellikler ise en az öneme sahip özellikler olarak görülmüştür.

3 EMOSTAR

Bu çalışmada EmoSTAR adlı yeni bir duygu veritabanını tanıtılacaktır. EmoSTAR televizyon ve internetteki kaynaklardan derlenerek hazırlanmış bir veri setidir. Nötr duygu içeren örnekleri bulmak (haber kanalları gibi) kolay olsa da diğer duyguların toplanmasında büyük zorluklar ortaya çıkmaktadır. Bazı durumlarda arka plandaki sesler ve müzik de sorun oluşturmaktadır. Yapay duygu elde ederken de katılımcıların doğal halleriyle konuşmaları kolay olmamaktadır.

EmoSTAR toplamda 393 olmak üzere kızgın, nötr, mutlu ve üzgün Türkçe ve İngilizce örneklerden oluşmaktadır. Nötr örnekler haber kanallarından, kızgın örnekler sinema ve dizilerden, üzgün örnekler internetteki videolardan, mutlu örnekler ise Oscar, Golden Globe gibi ödül törenlerinde ödül alan sanatçıların yaptığı konuşmalardan derlenmiştir. Nötr ve mutlu örnekler doğal konuşma içermektedir. Kızgın örnekler yapay duyguludurlar ve üzgün örnekler çok azı hariç doğal konuşmalardan oluşmaktadır. Örneklerin hepsi farklı cümleler içermektedir ve uzunlukları 2,2 ile 14,5 saniye arasında değişmektedir. Konuşmacı başına örnek sayısı ise 1 ile 29 arasındadır. Kategorik duygu etiketleme yazar tarafından sesli ve görsel değerlendirmeye videoların sahibi tarafından atanan üzgün ve kızgın etiketlerde göz önüne alınarak yapılmıştır. Veri setindeki örnek sayıları Çizelge 1'de verilmiştir.

Çizelge 1. EmoSTAR'da duygu sayıları (İ=İngilizce, T=Türkçe)

	Kızgın	Nötr	Mutlu	Üzgün
Erkek	33 İ - 30 T	35 İ - 34 T	45 İ	12 İ
Kadın	40 E	37 İ - 20 T	37 İ	51 İ - 19 T
Toplam=393	103	126	82	82

Üzerinde çalışacağımız diğer veritabanı ise Berlin Duygu Veritabanıdır (EmoDB). EmoDB Çizelge 2’de gösterildiği üzere, 7 duygu içeren 535 Almanca örnekten oluşmaktadır. Bu örneklerde 5 kısa ve 5 uzun cümle, 5 erkek ve 5 kadın tarafından seslendirilmiştir. Pan vd., EmoDB üstünde mutlu, nötr, üzgün sınıflandırmada % 95,1 başarı elde etmiştir. Mutlu, nötr, üzgün, sıkkin ve tiksinti sınıflandırmasında ise enerji ve prozodi özellikleri % 66, LPCMCC (Doğrusal Öngörü Kodlama Mel Kepstrum Katsayıları) % 70, her ikisi birlikte % 82 başarı sağlamıştır (Pan vd., 2012). Wu vd., EmoDB’de çok-sınıflı SVM ve 10-kat çapraz geçirme ile 7 duygu üstünde % 85.4 başarı elde etmiştir (Wu vd., 2010).

Çizelge 2. EmoDB’de duygu sayıları; (K:kızgın, N:nötr, M:mutlu, Ü:üzgün, S:sıkkin, T:tiksinti, Ko:korku)

	K	N	M	Ü	S	T	Ko
Erkek	60	39	27	25	35	11	36
Kadın	67	40	44	37	46	35	33
Toplam=535	127	79	71	62	81	46	69

4. DENEYSEL KURULUM

Denemeler, openSMILE ve Weka araçlarıyla gerçekleştirilmiştir (Eyben vd., 2009; Hall vd., 2009). Özellik çıkarma işleminde openSMILE ile beraber gelen Emobase ve Emo_large konfigürasyonlarında MFCC sayısı 12’den 24’e çıkarılmış ayrıca, Harmonic-to-Noise-Ratio özelliği eklenmiştir. Bu dosyalardaki özellik sayıları Çizelge 3’te gösterilmektedir.

Çizelge 3. Konfigürasyonlardaki özellik sayıları

Konfigürasyon	Özellik Sayısı
Emobase.conf	1482 (39 LLD + 39 delta)*19 fonksiyon
Emo_large.conf	8190 (70 LLD + 70 delta+ 70 delta-delta)*39 fonksiyon

Deneylerin tamamı 10-kat çapraz geçirmeye Weka’nın Naive Bayes (NB) ve SMO sınıflandırıcılarıyla gerçekleştirilmiştir. EmoDB ve EmoSTAR için ağırlıklı ortalama doğruluk oranları Çizelge 4, 5 ve 6’da verilmektedir.

Çizelge 4. EmoDB’de 7 duygu için sınıflandırma sonuçları

EmoDB	Emobase (988)	Emo_large (6669)
NB	57,00	70,46
SMO	87,85	87,28

Çizelge 5. EmoDB’de 4 duygu (Kızgın, Nötr, Mutlu, Üzgün) için sınıflandırma sonuçları

EmoDB	Emobase (1482)	Emo_large (8190)
NB	80,82	84,36
SMO	91,44	91,15

Çizelge 6. EmoSTAR’da 4 duygu (Kızgın, Nötr, Mutlu, Üzgün) için sınıflandırma sonuçları

EmoSTAR	Emobase (1482)	Emo_large (8190)
NB	83,20	86,00
SMO	95,92	96,69

5. VERİ SETLERİ ARASI ÇAPRAZ TESTLER

Duygu çıkarmadaki 20 yılın ardından tek veri setli çalışmalarda başarı oranı yönünden limitlere ulaşılmış gibi gözükmektedir. Setler arası çapraz testler araştırmacılar için üstesinden gelinmesi gereken yeni bir konu olarak karşımıza çıkmaktadır. Bu tür uygulamalarda tek veri setli uygulamalara göre başarı oranları oldukça düşmektedir. Çapraz testleri veri setlerimizden birini eğitim seti diğerini de test seti olarak kullanarak gerçekleştirdik. EmoSTAR 4 duygu içerdiği için testlerde EmoDB'den de Çizelge 7'de gösterildiği gibi 4 duygu kullanılmıştır.

Çizelge 7. EmoSTAR ve EmoDB eğitim ve test seti olarak çapraz test sonuçları

	EmoSTAR Eğitim EmoDB Test		EmoDB Eğitim EmoSTAR Test	
	NB	SMO	NB	SMO
Emobase	43,65	52,80	41,73	43,00
Emo_large	45,13	64,30	41,98	43,25

6. ÖZELLİK SEÇME

Bir veri setinde en iyi sonucu üreten optimal özellikleri bulmak araştırmacıların en fazla gayret sarf ettikleri konulardan birisidir. Gereğinden fazla özellik kullanmak sınıflandırıcıların başarı oranlarında negatif etki oluşturmaktadır. Bu çalışmada da Çizelge 4, 5 ve 6'da bu sorun kendini göstermiştir. Çizelge 5'te görüldüğü gibi özellik sayısının 1482'den 8190'a çıkmasına rağmen en başarılı sınıflandırıcı SMO'nun başarı oranında bir düşüş meydana gelmiştir. Bu sorunu çözmek amacıyla özellik seçicilerden yararlanılmaya karar verilmiştir. Özellik seçme için Information Gain, ChiSquared, Principal Components gibi özellik seçiciler, Ranker arama yöntemiyle; CfsSubSet özellik seçici ise Linear Forward Selection arama yöntemiyle kullanılmış ve sonuçlar EmoDB için Çizelge 8 ve 9'da, EmoSTAR için Çizelge 10 ve 11'de verilmiştir. EmoDB'de InfoG+Rank ve Chi+Rank % 88,41 başarı oranına 1074 özellikle ulaşılmıştır. Bu oran 1482 ve 8190 özellikle elde edilen oranın üstündedir.

Çizelge 8. EmoDB'de özellik seçme (Emobase)

EmoDB	Emobase (1482) ‡	
	NB (57,00)*	SMO (87,85)*
CfsSub+LFS	76,44	81,68
InfoG+Rank	69,90	88,41
Chi+Rank	69,90	88,41
PCA (145) †	46,91	74,01

* Özelliklerin tamamıyla elde edilen sonuçlar † Seçilen özellik sayıları ‡ Tüm özellik sayısı

Aynı özellik seçici, Emo_large konfigürasyonunda da daha az özellikle tüm özellik setinden daha iyi bir başarı oranı vermiştir. Emobase konfigürasyonunda, Naive Bayes sınıflandırıcının performansında büyük artış meydana gelmiştir. Bu sonuç aşırı özellik sayısının performans üstündeki negatif etkisinin bir kanıtıdır.

Çizelge 9. EmoDB’de özellik seçme (Emo_large)

EmoDB	Emo_large (8190)	
	NB (70,46)*	SMO (87,28)*
CfsSub+LFS (102) †	77,75	83,55
InfoG+Rank (6512) †	70,09	87,28
Chi+Rank (6512) †	69,90	87,47

EmoSTAR’da ise InfoG+Rank ve Chi+Rank % 97,20 başarı oranı ile en başarılı seçici olmuş ve Çizelge 10 ve 11’de görüldüğü gibi tüm özelliklerden daha iyi sonuç elde edilmiştir. Buna karşılık, PCA diğerlerine göre zayıf bir performans sergilemiştir. PCA, Emobase konfigürasyonunda 145 özellik seçmesine rağmen en az özellik seçen CfsSubSet seçicinin başarısının altında kalmıştır. PCA genellikle 150-200 özellik grubu önermekte ve bunları belli bir sıraya koymamaktadır. En iyi sonucu üreten özellik grubunu bulmak oldukça zor ve zahmetli bir iş olmaktadır. Bu çalışmada ilk önerilen grup seçilmiş ve en iyi grubu bulmak için bir çalışma yapılmamıştır.

Çizelge 10. EmoSTAR’da özellik seçme (Emobase)

EmoSTAR	Emobase (1482)	
	NB (83,20)*	SMO (95,92)*
CfsSub+LFS (75) †	87,27	94,65
InfoG+Rank (1236) †	83,96	96,18
Chi+Rank (1236) †	83,96	96,18
PCA (105) †	72,26	81,67

Çizelge 11. EmoSTAR’da özellik seçme (Emo_large)

EmoSTAR	Emo_large (8192)	
	NB (86,00)*	SMO (96,69)*
CfsSub+LFS (95) †	89,05	94,14
InfoG+Rank (6755)	85,49	97,20
Chi+Rank (6755) †	84,98	97,20

7. SONUÇLAR

Duygu, insanları diğer varlıklardan ve makinelerden ayıran en belirgin özelliklerdendir ve insanlar arasındaki iletişimde çok önemli bir yer tutmaktadır. Duygu tanıma çalışmalarının en önemli elemanlarından biri duygu veritabanlarıdır. Bu çalışmada EmoSTAR adı verilen yeni bir duygu veritabanı sunulmuş ve mevcut veritabanlarından biri olan emoDB ile çapraz testler gerçekleştirilmiştir.

Makalenin en önemli sonuçlarından biri setler arası çalışmalarda orta düzey başarı oranlarıdır. 8190 özellik kullanan Emo_large özellik seti, EmoSTAR veri setinin eğitim seti olarak kullanıldığı çapraz testlerde güçlü bir performans sergilemiştir. Ancak, EmoDB’de tek veri setli çalışmalarda Emobase özellik setinin gerisinde kalmıştır.

Ümit verici bir diğer sonuç ise, özellik seçme algoritmalarının çarpıcı performanslarıdır. Özellik seçiciler çok büyük özellik sayılarına rağmen, orijinal özellik setlerinden daha iyi başarı oranlarını yakalayabilmektedir.

İleriki çalışmalarda EmoSTAR veri setinin daha fazla duygu ve daha fazla örnek ile geliştirilmesi düşünülmektedir. Veri setleri arası çapraz testler de daha fazla veri seti ile gerçekleştirilecektir.

KAYNAKLAR

- Batliner A., Steidl S., Schüller B., Seppi D., Vogt T., Wagner J., Devillers L., Vidrascu L., Aharonson V., Kessous L. ve Amir N. (2009): "Whodunnit Searching for the Most Important Feature Types Signalling Emotion-Related User States in Speech: Appendix", Preprint submitted to Elsevier 24 January 2010, Computer Speech and Language, doi:10.1016/j.csl.2009.12.003.
- Bhargava M., Polzehl T. (2012): "Improving Automatic Emotion Recognition From Speech Using Rhythm and Temporal Feature", Proceedings of Icecit-2012 Published by Elsevier, s.139.
- Black A. W., Bunnell H. T., Dou Y., Kumar P., Metze F., Perry D., Polzehl T., Prahallad K., Steidl S. ve Vaughn C. (2011): "New Parameterization for Emotional Speech Synthesis", CSLP Proc., Johns Hopkins Summer Workshop, Baltimore.
- Chavhan Y., Dhore M. L., Yesaware P. (2010): "Speech Emotion Recognition Using Support Vector Machine", International Journal of Computer Applications (0975 - 8887), Cilt 1, No. 20.
- Eyben F., Wöllmer M., Schuller S. (2009): "openSMILE-The Munich Versatile and Fast Open-Source Audio Feature Extractor", In Proc. ACM Multimedia (MM), ACM, Florence, Italy, ACM, ISBN 978-1-60558-933-6, pp. 1459-1462, doi:10.1145/1873951.1874246.
- Eyben F., Batliner A., Schuller B., Seppi D., Steidl S. (2010): "Cross-Corpus Classification of Realistic Emotions – Some Pilot Experiments", In Proc. 7th Intern. Conf. on Language Resources and Evaluation (LREC 2010), Valletta, Elra, 2010.19.-21.05.2010.
- Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H. (2009): "The Weka data mining software: an update". Sigkdd Explor. Newsl. Cilt 11, s.10–18.
- He L. (2010): "Stress and Emotion Recognition in Natural Speech in the Work and Family Environments", PhD, Rmit University.
- Iida A., Campbell N., Higuchi F., Yasumura M. (2003): "A corpus-based speech synthesis system with emotion", Speech Communication, Cilt 40, s.161–187
- Mairesse F., Polifroni J., Di Fabrizio G. (2012): "Can Prosody Inform Sentiment Analysis? Experiments On Short Spoken Reviews", Nokia Research, ICASSP.
- Oflazoğlu C., Yildirim S. (2013): "Recognizing emotion from Turkish speech using acoustic features", Eurasip Journal on Audio, Speech, and Music Processing.
- Pan Y., Shen P., Shen L. (2012): "Speech Emotion Recognition Using Support Vector Machine", International Journal of Smart Home, Cilt 6, No. 2.
- Ramakrishnan S. (2012): "Recognition of Emotion from Speech: A Review", International Journal of Speech Technology, Cilt 15, Sayı 2, s.99-117.
- Schuller B., Batliner A., Seppi D., Steidl S., Vogt T., Wagner J., Devillers L., Vidrascu L., Amir N., Kessous L., Aharonson V. (2007): "The Relevance of Feature Type for the Automatic Recognition of Emotional User States: Low Level Descriptors and Functionals", Interspeech, Antwerp.
- Schuller B., Vlasenko B., Eyben F., Rigoll G., Wendemuth A. (2009): "Acoustic Emotion Recognition: A Benchmark Comparison of Performances", IEEE Workshop on Automatic Speech Recognition & Understanding, Asru 2009 Proc., Merano.

- Schuller B., Vlasenko B., Eyben F., Wollmer M., Stuhlsatz A., Wendemuth A., Rigoll G. (2010): "Cross-corpus acoustic emotion recognition: variances and strategies", IEEE Transactions on Affective Computing, Cilt 1, Sayı 2, s.119–131.
- Schuller B., Zhang Z., Weninger F., Rigoll G. (2011): "Selecting Training Data for Cross-Corpus Speech Emotion Recognition: Prototypicality vs. Generalization", Speech Processing Conference, Avios Proc., Telaviv.
- Shahzadi A., Ahmadyfard A., Yaghmaie K., Harimi A. (2013): "Recognition of Emotion In Speech Using Spectral Patterns", Malaysian Journal of Computer Science, Cilt 26, Sayı 2, s.143-158
- Wu S., Falk T. H., Chan W. (2010): "Automatic speech emotion recognition using modulation spectral features", Speech Communication, doi:10.1016/j.specom.2010.08.013.
- Wu D., Parsons D.T., Narayanan S. S. (2011): "Acoustic Feature Analysis in Speech Emotion Primitives Estimation", Interspeech 2011, 26-30, Makuhari, Chiba, Japan.
- Zhang Z., Weninger F., Wöllmer M., Schuller B. (2011): "Unsupervised Learning in Cross-Corpus Acoustic Emotion Recognition", IEEE Workshop on Automatic Speech Recognition & Understanding, Asru Proc., Waikoloa, Hawaii.