

# Duygu Tanımada Farklı Veri Setleri Arasında Bir Çalışma

Cevahir PARLAK<sup>1</sup> Banu DIRİ<sup>2</sup>

<sup>1,2</sup>Bilgisayar Mühendisliği Bölümü  
Elektrik-Elektronik Fakültesi  
Yıldız Teknik Üniversitesi, Beşiktaş, İSTANBUL

Eposta: cevahir.parlak@boun.edu.tr banu@ce.yildiz.edu.tr

## Özet

*Bu çalışmamızda 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. Özellik seçicilerin performansı da incelenmiştir. Özellik çıkarma openSMILE Emobase ve Emo\_large konfigürasyonlarında MFCC sayısı 12'den 24'e çıkartılarak ve Harmonic-to-Noise-Ratio (HNR) özellikleri eklenerek gerçekleştirilmiştir. Özellik seçme ve sınıflandırma ise Weka aracıyla yapılmıştır. EmoSTAR halen daha fazla duygu türü ve örnek için geliştirilme aşamasındadır.*

## 1. Giriş

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 yapabilmek ve karşısındakinin duygusunu anlayabilmek özelliği kazandırılmaya çalışılmaktadır. Böylece insan-makine etkileşimi daha çekici olabilmektedir. Duygu tanıma çalışmalarının temel öğelerinden birisi de hiç kuşkusuz duygu veritabanlarıdır. Ne var 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 tanımının uygulama alanları arasında çağrı merkezi uygulamaları, bilgisayar destekli öğrenme sistemleri, yalan makineleri, sesli eposta sistemleri ve oyunlar sayılabilir. Son yıllarda otomobillerde de sürücünün performansı ile duygu hali arasındaki ilişkiye gözlemlemek amacıyla duygu tanımadan yararlanılmaktadır [1]. Duygusal konuşma sentezi alanında çalışmalar da gün geçtikçe artmaktadır [2], [3].

## 2. İlgili Çalışmalar

Son yirmi yıl içinde duygu tanımadaki çalışmalar büyük ivme kazanmıştır. Pek çok veritabanı geliştirilmiş ve veritabanları üstünde sayısız çalışma gerçekleştirilmiştir [4], [5], [6]. Teknolojinin gelişmesiyle üstünde çalışılan duygu sayısı artmış ve çok yüksek başarı oranlarına ulaşılmıştır. Ancak bu çalışmalarda önemli sorunlardan biri veri setleri arasında uygulanan testlerde karşılaşılan düşük başarı oranlarıdır. Setler arası, karma setli ve çok setli çalışmalara [7-10] literatürde nispeten daha az rastlanılmaktadır ve bu alanda çözülmeyi bekleyen sorunlar bulunmaktadır.

Oflazoğlu ve Yıldırım [8] TURES (Turkish Emotional Speech Database) ve VAM (Vera am Mittag) veri setlerinde setler arası bir çalışma gerçekleştirmiş ve 3 boyutlu duygu uzayında %41.3 başarı oranı elde etmişlerdir.

Zhang vd. [9] 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.

Schüller vd. [11] EmoDB'nin test seti, AVIC, DES, eINTERFACE, SmartKOM ve SUSAS'ın eğitim seti olarak kullanıldığı çalışmalarında %35 ile %45 arasında başarı oranları elde etmişlerdir.

ABC, AVIC, DES, EmoDB, eNTERFACE, SAL, SUSAS ve VAM olmak üzere 8 veri setinin kullanıldığı bir çalışmada [12], VAM üstünde eğitim diğer 7 veri seti üstünde de test yapılarak iki boyutlu duygu uzayında aktivitede %67.7 başarı oranı yakalanmıştır. DES üstünde eğitim diğer 7 veri seti üstünde test yapılarak pozitiflik boyutunda %54.8 başarı oranı yakalanmıştır.

### 3. EmoSTAR

Bu çalışmamızda EmoSTAR adlı yeni bir duygu veritabanını tanıtacağız. 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 etmek amacıyla katılımcıların doğal halleriyle konuşmaları kolay olmamaktadır.

EmoSTAR 393 kızgın, nötr, mutlu ve üzgün Türkçe ve İngilizce örnekten 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. Ayrıntılı sayılar Tablo 1. de verilmiştir.

Üzerinde çalışacağımız diğer veritabanı ise Berlin Duygu Veritabanıdır (EmoDB). EmoDB Tablo 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 seslendirilmektedir. Pan [5] EmoDB üstünde mutlu, nötr, üzgün sınıflandırmada %95.1 başarı elde etmiştir. Mutlu, nötr, üzgün, sıkın ve tiksinti sınıflandırmasında ise enerji ve prozodi özellikleri %66, LPCMCC (Linear Prediction Coding Mel Cepstrum Coefficients) 70%, her ikisi birlikte %82 başarı sağlamıştır. Wu vd. [6] EmoDB’de çok-sınıflı

SVM ve 10-kat çapraz geçirme ile 7 duygu üstünde %85.4 başarı oranına ulaşmışlardır.

Tablo 1. EmoSTAR’da duygu sayıları (İ=İngilizce, T=Türkçe).

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

Tablo 2. EmoDB’de duygu sayıları.

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

### 4. Deneysel Kurulum

Deneyler openSMILE [11] ve Weka [12] araçlarıyla gerçekleştirilmiştir. 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 Tablo3’te gösterilmektedir.

Tablo 3. Konfigürasyonlardaki özellik sayıları.

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

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

Tablo 4. EmoDB’de 7 duygu için sınıflandırma sonuçları.

EmoDB	Emobase (988)	Emo_large (6669)
<b>NB</b>	57.00	70.46
<b>SMO</b>	<b>87.85</b>	87.28

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

EmoDB	Emobase (1482)	Emo_large (8190)
NB	80.82	84.36
SMO	<b>91.44</b>	91.15

Tablo 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	<b>96.69</b>

## 5. Veri Setleri Arası 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 Tablo 7’de gösterildiği gibi 4 duygu kullanılmıştır.

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

	EmoSTAR Eğitim		EmoDB Eğitim	
	EmoDB Test		EmoSTAR Test	
	NB	SMO	NB	SMO
Emobase	43.65	<b>52.80</b>	41.73	43.00
Emo_large	45.13	<b>64.30</b>	41.98	43.25

## 6. Özellik 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 Tablo 4, 5 ve 6’da bu sorun kendini göstermiştir. Tablo 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 yararlanılmaktadır. Özellik seçme için Information

Gain, ChiSquared, Principal Components özellik seçiciler Ranker arama yöntemiyle, CfsSubSet özellik seçici ise Linear Forward Selection arama yöntemiyle kullanılmış ve sonuçlar EmoDB için Tablo 8 ve 9’da, EmoSTAR için Tablo 10 ve 11’de verilmiştir. EmoDB’de InfoG+Rank ve Chi+Rank %88.41 başarı oranına 1074 özellekle ulaşmıştır. Bu oran 1482 ve 8190 özellekle elde edilen oranın üstündedir.

Tablo 8. EmoDB’de özellik seçme (Emobase).

EmoDB	Emobase (1482) ‡	
	NB (57.00)*	SMO (87.85)*
CfsSub+LFS (55) †	76.44	81.68
InfoG+Rank (1074)	69.90	<b>88.41</b>
Chi+Rank (1074) †	69.90	<b>88.41</b>
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 özellekle tüm özellik setinden daha iyi bir başarı oranı elde etmiştir. Emobase konfigürasyonunda Naive Bayes sınıflandırıcının performansında büyük artış meydana gelmiştir. Bu sonuç aşırı özellik sayısının performans üstündeki negatif etkisinin bir kanıtıdır.

Tablo 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	<b>87.47</b>

EmoSTAR’da ise InfoG+Rank ve Chi+Rank %97.20 başarı oranı ile en başarılı seçici olmuş ve Tablo 10 ve Tablo 11’de görüldüğü gibi tüm özelliklerden daha iyi sonuç elde etmiştir. PCA diğerlerine göre zayıf bir performans sergilemiştir. PCA, Emobase konfigürasyonunda 145 özellik seçmesine rağmen en az özellik seçen CfsSubSet seçicinin 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.

Tablo 10. EmoSTAR'da özellik seçme (Emobase).

EmoSTAR	Emobase (1482)	
	NB (83.20)*	SMO (95.92)*
<b>CfsSub+LFS</b> (75) †	87.27	94.65
<b>InfoG+Rank</b> (1236) †	83.96	<b>96.18</b>
<b>Chi+Rank</b> (1236) †	83.96	<b>96.18</b>
<b>PCA</b> (105) †	72.26	81.67

Tablo 11. EmoSTAR'da özellik seçme (Emo\_large).

EmoSTAR	Emo_large (8192)	
	NB (86.00)*	SMO (96.69)*
<b>CfsSub+LFS</b> (95) †	89.05	94.14
<b>InfoG+Rank</b> (6755) †	85.49	<b>97.20</b>
<b>Chi+Rank</b> (6755) †	84.98	<b>97.20</b>

## 7. Sonuçlar

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 yakalayabilmiştir.

İleriki çalışmalarda EmoSTAR veri seti daha fazla duygu ve daha fazla örnekle geliştirilecektir. Veri setleri arası çapraz testler de daha fazla veri seti ile gerçekleştirilebilir.

## 8. Kaynaklar

[1] Ramakrishnan, S., "Recognition of Emotion from Speech: A Review", International Journal of Speech Technology, v:15, Issue 2, pp 99-117, 2012.

[2] Black, A. W., Bunnell, H.T., Dou, Y., Kumar, P., Metzger, F., Perry, D., Polzehl, T., Prahallad, K., Steidl, S. ve Vaughn, C., "New Parameterization for Emotional Speech Synthesis", CSLP Proc., Johns Hopkins Summer Workshop, Baltimore, 2011.

[3] Iida, A., Campbell, N., Higuchi, F. ve Yasumura M., "A corpus-based speech synthesis system with emotion", Speech Communication 40 161–187, 2003.

[4] Pan, Y., Shen, P. ve Shen, L., "Speech Emotion Recognition Using Support Vector Machine", International Journal of Smart Home, v:6, no. 2, Apr. 2012.

[5] Schuller, B., Vlasenko, B., Eyben, F., Rigoll, G. ve Wendemuth, A., "Acoustic Emotion Recognition: A Benchmark Comparison of Performances", IEEE Workshop on Automatic Speech Recognition & Understanding, ASRU 2009 Proc., Merano, Dec. 2009.

[6] Wu, S., Falk T.H. ve Chan W., "Automatic speech emotion recognition using modulation spectral features", Speech Communication 2010, doi:10.1016/j.specom.2010.08.013.

[7] Oflazoglu C. ve Yildirim S., "Recognizing emotion from Turkish speech using acoustic features", EURASIP Journal on Audio, Speech, and Music Processing, 2013.

[8] Zhang, Z., Wenginger, F., Wöllmer, M. ve Schuller, B., "Unsupervised Learning in Cross-Corpus Acoustic Emotion Recognition", IEEE Workshop on Automatic Speech Recognition & Understanding, ASRU Proc., Waikoloa, Hawaii, Dec. 2011.

[9] Schuller, B., Vlasenko, B., Eyben, F., Wollmer, M., Stuhlsatz, A., Wendemuth, A. ve Rigoll, G., "Cross-corpus acoustic emotion recognition: variances and strategies", IEEE Transactions on Affective Computing 1(2), 119–131, July-December 2010.

[10] Schuller, B., Zhang, Z., Wenginger, F. ve Rigoll, G., "Selecting Training Data for Cross-Corpus Speech Emotion Recognition: Prototypicality vs. Generalization", 2011 Speech Processing Conference, AVIOS Proc., Telaviv, June 2011.

[11] Eyben, F., Wöllmer, M. ve 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, October 2010. doi:10.1145/1873951.1874246

[12] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., (2009). "The WEKA data mining software: an update". SIGKDD Explor. Newsl. 11, 10–18 (2009).