

3-Boyutlu Duygu Uzakında Konuşma Sinyalinden Duygu Tahmini ile Değerlendiricilerin Analizi

Analysis of Evaluators based on Emotion Estimation from Speech in 3 Dimensional Emotional Space

Çağlar OFLAZOĞLU, Serdar YILDIRIM, Esen YILDIRIM, Yakup KUTLU

Bilgisayar Mühendisliği Bölümü
Mustafa Kemal Üniversitesi

[caglaroflazoglu,serdar,yildirim,ykutu]@mku.edu.tr

Özet

Konuşma sinyalinden duygu tanıma için öncelikle modellemede kullanılacak konuşma sinyallerinin duygu içeriklerinin belirlenmesi gerekmektedir. Bu işlem için birçok değerlendiriciyle dinleme testleri yapılmakta ve değerlendiricilerden elde edilen sonuçlar yorumlanarak konuşma sinyallerinin duygu sınıfı belirlenmektedir. 3-boyutlu duygu uzayı değerlendirilmede değerlendiricilerin her bir konuşma için verdikleri değerlerin ortalaması veya ortanca değeri alınarak duygu değerleri belirlenmektedir. Bu çalışmada 3-boyutlu duygu uzayında değerlendiricilerin değerlendirmeleri ayrı ayrı kullanılarak sınıflandırma çalışması yapılmış ve elde edilen performanslar ortalama ve ortanca değerler kullanılarak elde edilen sonuçlarla karşılaştırılmıştır. Sonuçlar incelendiğinde genel olarak bütün boyutlar için ortalamanın en iyi performansı verdiği görülmektedir.

Abstract

To recognize emotion from speech, emotional content of speech signals which are used for modelling should be determined first. Emotion classes are determined from evaluations made by numerous evaluators via listening tests. In assessment of 3-dimensional emotion space, average or median of ratings from evaluators for each speech is used. In this study, we compared estimation performances of individual evaluators' ratings to performances obtained using average and median values. Results show that best results are obtained using average values for all emotion dimensions.

1. Giriş

Teknolojideki gelişmelerle birlikte insan-makine etkileşimini daha doğal hale getirecek sistemlerin tasarlanması önem kazanmıştır. Biyometrik işaretlerden elde edilen özelliklerin kullanılması ile araştırmacıların duygu tanıma üzerine yaptıkları çalışmalar dikkat çekici seviyede artmıştır.

Duygu içeriğinin ifade edilmesinde geleneksel yaklaşım duygusal içeriklerin ayrık duygu kategorileri (üzgün, kızgın, mutlu vb.) olarak kabul edilmesidir. Duygusal içerikleri ifade etmede bu yöntem tercih edildiğinde duygunun herhangi bir duygu kategorisine dahil edilememesi ya da birden fazla sınıfı içerebilmesi durumu ortaya çıkabilmektedir. Psikoloji alanında yapılan çalışmalarla birlikte duygusal içeriklerin klasik kategoriler dışında çok boyutlu bir uzayda nokta veya bölge olarak

temsil edilebileceği ortaya konmuştur [1, 2, 3, 4]. Duygusal durumların çok boyutlu bir uzay üzerinde temsil edilebilmesinin duyguyu daha doğru şekilde tanımladığı düşünülmektedir. Literatürde Yu ve ark.[5] duygusal uzayı pozitiften negatife (*değerlik*) ve uyumlu/durgun olmadan heyecanlıya (*uyarılma*) doğru bir skalada kabul ederek yaptıkları çalışmada %54 ile %69 arasında bir tahmin başarı elde etmişlerdir. Vidrascu ve Devillers[6] 10 saatlik acil servis kayıtlarını kullanarak yaptıkları çalışmada pozitif ve negatif değerleri kullanarak %82 başarı elde etmişlerdir. Wu ve ark.[7] VAM veritabanını kullanarak çıkardıkları öznitelik vektörleri ile duygu durumlarını 3-boyutlu uzay üzerinde modellemeye çalışmışlardır.

Literatürde *değerlik* (Valence), *aktivasyon* (Activation) ve *baskınlık* (Dominance) duygu düzlemleri duygu tanıma çalışmalarında sıklıkla kullanılmaktadır. *Değerlik*, pozitiflikten negatifliğe, *aktivasyon* düşük uyarımdan yüksek uyarıma ve *baskınlık* ise konuşmacının dinleyici üzerindeki etkisinin zayıflıktan güçlüye doğru olduğu bir skalayı temsil etmektedir. Duygu durumunun tespitinde değerlendiriciler çok boyutlu olan bu düzlem üzerinde yaptıkları işaretlemelerle duygu durumunu belirler. Fakat duygusal durumun algısı kişiden kişiye göre değişen bir kavramdır. Bu nedenle konuşma işaretine ait değerlendirici sonuçları duygu uzayında bir noktadan çok bir bölge oluşturabilir. Yapılan benzer çalışmalar incelendiğinde değerlendiricilerden elde edilen çok boyutlu duygu uzay değerlendirmelerinin tek bir doğru sonuca dönüştürülmesi için ortalama kullanımının yaygın olduğu görülmektedir. Mover ve diğerleri yaptıkları çalışmada ortalama değer kullanmanın ayrı ayrı değerlendirici değerlendirmelerini kullanmaya göre çok daha iyi sınıflandırma performansı verdiğini göstermişlerdir [8]. Yapılan çalışmada sadece iki değerlendirici bulunmaktadır. Bu çalışmada ise 27 değerlendiricinin değerlendirmesine göre analizler 3-boyutlu duygu uzayı değerlerinin tahmini için incelemiştir.

Yapılan bu çalışmada değerlendiricilerin bireysel sonuçları ile tüm değerlendiricilere ait ortanca ve ortalama değerlendirme sonuçları kullanılarak çok boyutlu düzlemdeki duygusal durum tahmin performansları analiz edilecektir. Çalışmada değerlendiricilerin her birinin vermiş olduğu değerlendirme sonuçlarının doğru olduğu varsayımı ile 27 adet ve bu değerlerin ortalama ile ortanca değerlerine ait 2 adet olmak üzere toplam 29 adet veriseti kullanılmıştır.

Bu bildirinin devamında; Bölüm 2'de kullanılacak veritabanı hakkındaki bilgileri, Bölüm 3'te öznitelik çıkarımı ve seçim yöntemlerini, Bölüm 4'te kullanılan sınıflandırıcıyı,

Çizelge 1. Değerlendirici sonuçlarının ortalama ve ortanca değerler ile korelasyonuna ait sonuçlar

		<i>Değerlik</i>	<i>Aktivasyon</i>	<i>Baskınlık</i>
Ortalama Değerler	Ortalama	0.65	0.61	0.56
	Standart Sapma	0.08	0.10	0.12
	Maksimum	0.75	0.76	0.70
	Minimum	0.44	0.43	0.15
Ortanca Değerler	Ortalama	0.56	0.53	0.47
	Standart Sapma	0.07	0.11	0.12
	Maksimum	0.64	0.67	0.61
	Minimum	0.34	0.10	0.35

Bölüm 5’te elde edilen sonuçlar için değerlendirme yöntemlerini, Bölüm 6’da ise uygulanan yöntem hakkındaki bilgileri bulabilirsiniz. Son bölümde ise elde edilen sonuçların kısa bir irdelemesi verilmiştir.

2. Veritabanı

Yapılan çalışmada Türkçe Duygusal Konuşma veritabanı (TURES)[9] kullanılmıştır. Veritabanında Türkçe filmlerden elde edilmiş toplam 5100 adet konuşma işareti bulunmaktadır. Bu konuşma işaretleri 27 değerlendirici tarafından ayırık duygu kategorileri(kızgın, mutlu, üzgün, nötr vb.) ve 3-boyutlu duygu uzayında (değerlik, aktivasyon ve baskınlık) değerlendirilerek duygu içerikleri belirlenmeye çalışılmıştır. 3-boyutlu duygu uzayında duygunun yerinin belirlenmesi için Bradley ve Lang’ın geliştirdiği Self-Assessment Manikin (SAM) görselleri kullanılmıştır [10]. 3-boyutlu duygu uzayı değerlendirilmesi [-1 1] aralığında 5-adımlı skala kullanılarak yapılmıştır.

Kategorik değerlendirmede konuşma işaretlerinin duygu içeriği olarak çoğunluğun hem fikir olduğu duygu sınıfı seçilmiştir. Konuşma işaretlerinin 3-boyutlu duygu uzayındaki yeri ise değerlendiricilerin verdikleri değerlerin ortalaması alınarak elde edilmiştir. Değerlendiricilerin ortalama ve ortanca değerlerle uyumunu tespit edebilmek için değerlendiricilerin ortalama ve ortanca değerler arasındaki korelasyon sonuçlarının ortalamaları, standart sapmaları, maksimum ve minimum değerleri hesaplanarak Çizelge 1’de sunulmuştur.

3. Öznitelik Çıkarımı ve Seçimi

Çalışmada konuşma işaretlerinden elde edilen 1582 adet öznitelik openSMILE [11] yazılımı kullanılarak elde edilmiştir. Öznitelik vektörleri düşük seviye tanımlayıcı (LLDs) olarak tanımlanan temel frekans(F0), Mel frekans kepstrum katsayıları (MFKK), seslilik olasılığı, frekans ve genlik düzensizlikleri, darbe kod modülasyonu (DKM) sıfır geçiş oranı, DKM ortalama enerji karekökleri gibi özellikler ve bunların türevlerinden elde edilen istatistikleri içermektedir.

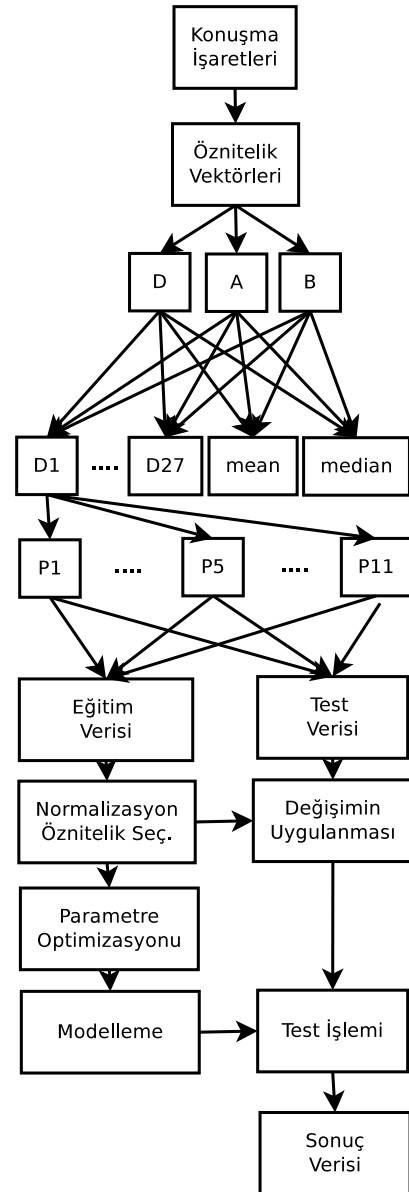
Çalışmada duyguları ayırt edici en uygun öznitelikleri belirlemek ve öznitelik vektör boyutunu azaltmak için öznitelik seçim işlemi gerçekleştirilmiştir. Öznitelik seçimi için ilinti tabanlı öznitelik seçim algoritması(CFS) tercih edilmiştir [12]. Bu yöntem birbiri ile ilintisiz ve sınıf ile yüksek ilintili öznitelik alt kümesini bulmayı amaçlayan bir algoritmadır(1).

S_k : k özniteliğini içeren S öznitelik altkümesi

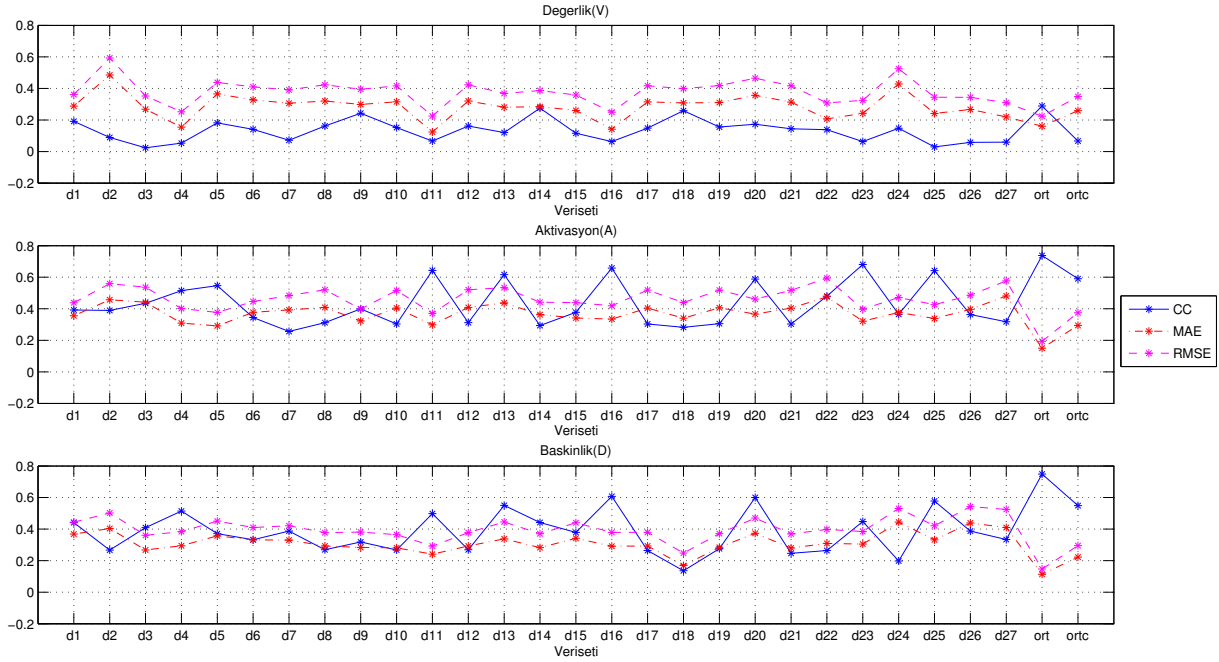
$\bar{r}_{cf}$: öznitelik-sınıf korelasyon değerleri ortalaması

$\bar{r}_{ff}$: öznitelik-öznitelik korelasyon değerleri ortalaması

$$Merit_{S_k} = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}}. \quad (1)$$



Şekil 1. Çalışmanın blok diyagramı (D:Değerlik, A:Aktivasyon, B:Baskınlık)



Şekil 2. Korelasyon Katsayısı(CC), Mutlak ortalama hata(MAE) ve Ortalama karesel hatanın karekökü(RMSE) sonuçlarının değerlendiricilere göre dağılımı[d1-d27] ile duygu uzayı değerlendirme sonuçlarının her bir konuşma işareti için ortalama(ort) ve ortanca(ortc) değerlerinin doğru olduğu varsayımı kullanılarak elde edilen sonuç dağılımları

4. Sınıflandırıcı

Çalışmada destek vektör regresyonu (ϵ - DVR) kullanılmıştır. ϵ - DVR için LibSVM kütüphanesi tercih edilmiştir [13, 14]. DVR modellemede çekirdek fonksiyon olarak radyal tabanlı çekirdek fonksiyonu kullanılmıştır. Bu yöntemde model parametrelerinin seçimi regresyon başarısı üzerinde etkilidir. En uygun model parametresi ϵ ve çekirdek parametresi γ 'nın belirlenmesinde LibSVM kütüphanesine ait ızgara-arama algoritması beş çapraz doğrulama ile kullanılarak elde edilmiştir.

5. Değerlendirme

Çalışmada sonuçlar 11-çapraz doğrulama yöntemi kullanılarak elde edilmiştir. Verisetindeki konuşma işaretleri rastgele 11 parçaya ayrılarak 10 parça eğitim ve bir parça test için kullanılmıştır. Her parça test için kullanılarak işlem tekrarlanmıştır. Aynı konuşmacıya ait konuşma işaretlerinin hem eğitim hem de test verisinde bulunmamasına dikkate edilecek çalışmanın konuşmacı bağımsız olması sağlanmıştır.

Sonuçların doğruluk oranlarını hesaplamak amacıyla yapılan çalışmalarda Pearson korelasyon katsayıları(CC) için (2), Ortalama mutlak hata(MAE) için (3), Ortalama karesel hatanın karekökü(RMSE) için ise (4) nolu eşitlik kullanılmıştır. Korelasyon değerinin 0.5-1.0 arasında olması yüksek derece, 0.30-0.49 arası orta, 0.10-0.29 arasında olması ise ilişkinin düşük derecede olduğu anlamına gelmektedir.

$$CC = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (2)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - y_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (4)$$

6. Gerçekleştirim

Konuşma işaretlerine ait öznitelik vektörlerinin elde edilmesinin ardından değerlendiricilerin değerlik, aktivasyon ve baskınlık düzlemlerinde her bir konuşma işareti için yaptıkları değerlendirmelere ait ortalama ve ortanca değerler hesaplanmıştır (Şekil 1).

Çalışmada değerlendiricilere ait değerlendirme sonuçlarının doğru olarak kabul edildiği 27 adet veriseti, bu değerlendirme sonuçlarının her bir konuşma işareti için ortalama ve ortanca değerlerinin hesaplanması sonucu 2 adet veriseti olmak üzere her bir duygu uzayı için toplam 29 adet veriseti elde edilmiştir (29 veriseti*3-boyut).

Her bir duygu uzayı ve değerlendirici için oluşturulan öznitelik vektörleri 11 grup olarak tanımlanmış olan konuşmacı listesine göre eğitim ve test verisi olarak parçalara ayrılmıştır (toplam $(11*2)*(29*3)$ adet vektör dosyası). Böylece aynı konuşmacılara ait konuşma işaretlerinin hem eğitim hem de test verisi içerisinde aynı anda yer almasının önüne geçilmiştir.

Eğitim ve test verilerinin elde edilmesinin ardından normalizasyon işlemi gerçekleştirilmiştir. Öznitelikler [-1:1] değer aralığına normalize edilmiştir. Eğitim verisi normalizasyonu sonucu oluşan parametre dosyası kullanılarak test verisi de normalize edilmiştir. Normalizasyonun ardından eğitim verisine CFS öznitelik seçim algoritması uygulanmış ve öznitelik-

Çizelge 2. Değerlendirici verisetlerinden elde edilen minimum, maksimum ve ortalama uyum sonuçları ile değerlendiricilere ait duygu uzay verilerinin ortalama ve ortanca değerinin modellenmesi ile elde edilen sonuçlar (CC: Korelasyon Katsayısı, MAE: Ortalama Mutlak Hata, RMSE: Ortalama Karesel Hatanın Karekökü)

		Değerlendirici Veriseti[1-27]			Tümü	
		Bireysel			Ortalama Değerler	Ortanca Değerler
		Minimum	Maksimum	Ortalama	Sonuç	Sonuç
Değerlik	CC	0.02	0.28	0.13	0.29	0.07
	MAE	0.12	0.49	0.29	0.16	0.26
	RMSE	0.23	0.59	0.38	0.22	0.35
Aktivasyon	CC	0.26	0.68	0.42	0.74	0.59
	MAE	0.29	0.48	0.38	0.15	0.30
	RMSE	0.37	0.60	0.47	0.19	0.37
Baskınlık	CC	0.14	0.60	0.37	0.75	0.55
	MAE	0.17	0.44	0.32	0.11	0.22
	RMSE	0.25	0.54	0.41	0.15	0.30

ler azaltılmıştır. Eğitim verileri arasından çıkarılan öznitelikler test verisinden de çıkarılmıştır.

7. Sonuçlar

Yapılan bu çalışmada üç ayrı duygu uzayı ve her bir veriseti için elde edilen sonuçlar Şekil 2 ve Çizelge 2’de verilmiştir. d1 - d27 olarak gösterilen verisetleri değerlendiricilerin her birinin duygu uzayında gerçekleştirmiş oldukları değerlendirme sonuçlarını, *ort* ve *ortc* verisetleri ise değerlendiricilerin her konuşma işareti için verdikleri sonuçların sırasıyla ortalamaları ve ortanca değerlerinin hesaplanması ile elde edilmiştir. Çalışmada her veriseti doğru sonuç olarak kabul edilip modellenmiş ve tahmin performansları elde edilmiştir.

Şekil 2’de verilen sonuçlar incelendiğinde bütün boyutlar için ortalama (*ort*) isimli verisetinin en iyi performansı verdiği görülmektedir. Ortanca (*ortc*) verisetini kullanmanın ise değerlik boyutu için uygun olmadığı ortaya konmuştur. *ortc* veriseti kullanılarak aktivasyon ve baskınlık boyutları için iyi sonuçlar elde edilmiş olsa da *ort* veriseti ile karşılaştırıldığında sonuçların düşük olduğu görülmektedir. Yapılan bu çalışma ile TurES veritabanının 3-boyutlu duygu analizi çalışmalarında, değerlendiricilerin ortalamalarından elde edilen verisetini kullanmanın en doğru seçim olacağı ortaya konmuştur.

8. Kaynaklar

- [1] Kehrein, R., 2002. The prosody of authentic emotions. In: Proc. Speech Prosody Conference, pp. 423-426.
- [2] Roddy Cowie and Randolph R. Cornelius. 2003. Describing the emotional states that are expressed in speech. Speech Commun. 40, 1-2 (April 2003), 5-32
- [3] H. Schlosberg, Three dimensions of emotion, Psychological Review, vol. 61, no. 2, pp. 81-88, 1954.
- [4] J. A. Russell, "Core affect and the psychological construction of emotion", Psychological Review, vol. 110, pp. 145-172, 2003.
- [5] Yu, C., Aoki, P.M., Woodruff, A., 2004. Detecting user engagement in everyday conversations. In: Proc. 8th Internat. Conf. on Spoken Language Processing (ICSLP), Vol. 2, Jeju Island, Korea, pp. 1329-1332.
- [6] Vidrascu, L., Devillers, L., 2005. Detection of real-life emotions in call centers. In: Proc. Eurospeech, pp. 1841-1844.
- [7] Wu, Dongrui and Parsons, Thomas D. and Narayanan, Shrikanth S.. Acoustic feature analysis in speech emotion primitives estimation. INTERSPEECH. editor(s) Kobayashi, Takao and Hirose, Keikichi and Nakamura, Satoshi. 785-788, ISCA, Year 2010.
- [8] Mower E., and Mataric M. J., and Narayanan S. Evaluating Evaluators: A Case Study in Understanding the Benefits and Pitfalls of Multi-Evaluator Modeling, In Proceedings of Interspeech 2009, 2009.
- [9] Oflazoglu, C.; Yildirim, S., "Turkish emotional speech database," Signal Processing and Communications Applications (SIU), 2011 IEEE 19th Conference on , vol., no., pp.1153,1156, 20-22 April 2011
- [10] M Bradley, P. Lang, Measuring emotion: The self-assessment manikin and the semantic differential. Journal of Behavioral Therapy and Experimental Psychiatry, 25, pp. 49-59, 1994.
- [11] F Eyben, M Wöllmer, B Schuller, in international conference on multimedia. openSMILE: the Munich versatile and fast open-source audio feature extractor (ACM Firenze, Italy, 25-29 Oct 2010)
- [12] Hall, M. A., Correlation-based feature selection for machine learning, Doctoral Dissertation, The University of Waikato, Department of Computer Science, 1999.
- [13] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011.
- [14] AJ Smola, B Schölkopf, A tutorial on support vector regression. Stat. Comput. 14(3), 199-222 (2004)