Konuşma Sentezinde Artan Doğallık için Boyut Azaltma Seçimi

Dimension Reduction Selection for Increasing Naturalness in Speech Synthesis

Mehmet Ali KUTLUGÜN
Bilgisayar Bilimleri ve Mühendisliği
İstanbul Sabahatin Zaim Üniversitesi
İstanbul , Türkiye
mehmet.kutlugun@std.izu.edu.tr

Yahya ŞİRİN
Bilgisayar Bilimleri ve Mühendisliği
İstanbul Sabahatin Zaim Üniversitesi
İstanbul , Türkiye
yahya.sirin@izu.edu.tr

***Özetçe*—Ses sentezleme sistemleri kullanıcı ile bilgisayar etkileşiminde insanlara büyük kolaylıklar sağlayan sistemlerdir. Bu sistemler insan konuşmasını yapay olarak işleyebildikleri gibi yazılı metinleri de okuyabilirler. Bu çalışmada monoton, robotik bir metin seslendirme yerine farklı metin türlerinin kendi kategorilerine göre farklı ses biçimleri şeklinde seslendirilmesi amaçlanmıştır. İnsan doğasına daha uygun olan bu yöntem için öncelikle ham metinler, metin önişleme aşamasından geçirilerek sınıflandırma işlemi yapılmıştır. İşlenen verilerden elde edilen bilgiler ile hangi metin türünün hangi ses tonlamaları ile seslendirileceğine karar verilmiştir. Kendi kategorisine uygun olmayan metinlerin seslendirilmesi, tonlamalarında farklılaştırma yapılarak sağlanmıştır. Böylece yanlış sınıflandırılan dokümanlar ses dosyalarındaki tonlama farklılıkları ile açık bir şekilde ayırt edilebilmiştir.**

Anahtar Kelimeler — ses sentezleme; makine öğrenmesi; sınıflandırma

***Abstract*— Speech synthesis systems are systems that provide great convenience to people on user-computer interaction. These systems can read written text as well as they can process human speech artificially. In this study instead of a monotone robotic speech synthesis, the idea is that it is better suited to human nature to have different types of texts sound differently in their own categories. For this process which is more suitable for human nature, the raw texts are first classified by passing through the text preprocessing step, The audio files are obtained by determining which audio texts are to be sounded with this tone as a result of this classification. Text speech that is not appropriate for its category is provided by differentiating its tonalities. Thus, misclassified documents can be clearly distinguished from tonal differences in audio files.**

Keywords — speech synthesis; machine learning; classification.

I. GİRİŞ

A. Ses Seslendirme

Bilgisayarların konuşması ve konuşulanları algılaması insan-bilgisayar etkileşiminin vazgeçilmez bir öğesidir [1].

978-1-5090-6494-6/17/$31.00 ©2017 IEEE

Bilgisayar bilimleri, elektronik mühendisliği, ses mühendisliği gibi farklı alanlar bilgisayar konuşmasını gerçekleştirebilmek için çalışan başlıca alanlardandır. Bu konuda iki temel çalışma alanı vardır. Bunlar; Metinden konuşma sentezleme ve konuşma tanıma sistemleridir.

Konuşma tanıma sistemleri, söylenilen herhangi doğal dili sayısal metin formatına çevirebilmekte ve dönüştürülen metinler içinde aramalar yaparak çeşitli örüntüler elde edebilmektedir. Bu sayede konuşulanları bilgisayar tarafından algılayıp çeşitli işlemleri gerçekleştirebilmektedirler.

Metinden Konuşma Sentezleme (MKS) ise, yazılı bir metnin elektronik ortamda ses sinyallerine dönüştürülme işlemidir. Bu metin kaynağı bir metin belgesi veya elektronik kitap olabileceği gibi bir web sayfası da olabilir. İdeal bir MKS sisteminin amacı insanın okuyabildiği her metni doğal insan sesi kalitesinde işleyebilmesidir. Başarımı yüksek bir MKS sistemi sayıları okuyabilmeli, kısaltmaları uygun formatta seslendirebilmeli ve bir kelimenin farklı yazımlarını ayırt edebilmelidir [1].

Bu alandaki ilk araştırma 1779 yılında Rus profesör Christian Kratzenstein tarafından yapılmıştır. 1791 yılında, Wolfgang von Kempelen bir makine geliştirmiş ve bazı sesleri bu makine ile elde etmeyi kısmen başarmıştır. 1800'lü yıllarda Charles Wheatstone, Kempelen'in cihazını geliştirerek daha iyi sonuçlar elde etmiştir[2]. Zaman içerisinde bu çalışmalar daha da geliştirilerek günümüze kadar devam etmiştir.

Başta İngilizce olmak üzere Türkçe dahil birçok dil için hazırlanmış MKS sistemleri ticari olarak son kullanıcıya sunulmaktadır. Sunulan ticari sistemler dışında Türkçe Metinden Konuşma Sentezleme (TMKS) alanında da akademik çalışmaların yapıldığı görülmektedir[3,4,5].

Ses sentezleyiciler genellikle yapay (robotik) sesler üretirler. Bu mekanik veya robotik ses insan sesinden kolaylıkla ayırt edilebilir. Bazı şartlar altında bu robotik ses tercih edilebilir fakat çoğu zaman sentezleyiciden gelen sesin

daha kolay anlaşılabilir ve dinlenebilir olabilmesi için insan sesine benzemesi tercih edilir. Konuşma sentezleyicilerin kalitesinin değerlendirmesi yapılırken iki önemli faktör üzerinde durulur. Bunlar anlaşılabilirlik ve doğallıktır. Anlaşılabilirlik sentezlenen konuşmanın kullanıcılar tarafından güvenli olarak anlaşılmasının göstergesidir. Doğallık ise sesin ne kadar insan sesine yaklaşabildiğiyle ve kullanıcılara bir insanla konuşuyormuş hissi vermesiyle alakalıdır [6].

Sentezlenen konuşmanın doğallığı sistemin tonlama, vurgulama, ritim, ölçü tekniği gibi özelliklerine bağlıdır. Ayrıca harflerin okunma süreleri, kullanılma sayıları ve metnin karmaşıklığı gibi ölçütler de sentezlenen konuşmanın doğallığını etkilemektedir[4]. Canal ve arkadaşları, Türkçe metinden konuşma sentezleme konusunda doğallığın artırılmasına yönelik yaptıkları çalışmalarda çeşitli çözüm yöntemleri ile insan sesine yakın bir metinden konuşma sentezleme sistemi geliştirilmişlerdir.

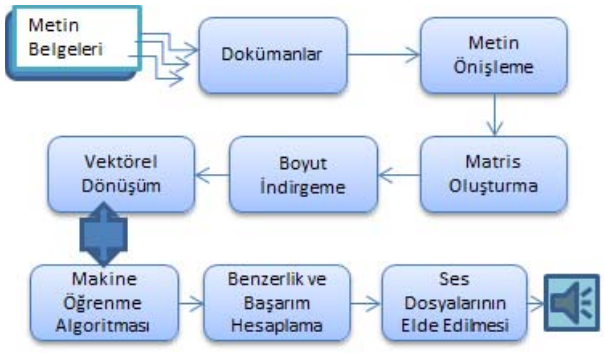
II. SUNULAN ÇALIŞMA

A. Amaç

Düz bir metinden en uygun ses biçiminin elde edilmesi için metnin doğru metotlarla işlenmesi ve yorumlanması gerekmektedir. Bunun için metinler önişleme sürecine tabi tutularak bazı hesaplamalar yapabilmek için sayısal değerlere dönüştürülmüştür. Bu sayısal (vektörel) değerler makine öğrenme algoritmaları yardımıyla başarımı en yüksek seviyede sınıflandırılmıştır. Sınıflandırılan bu belgelerin kendi sınıfına ait ses dosyaları ile insan doğasına en uygun seslendirme yapılması hedeflenmiştir. Bu işlemler için k-en yakın komşu (K-NN) algoritmasından yararlanılmıştır.

K-En Yakın Komşu modelinde bir elemanın sınıfını belirlemek için önceden sınıfları belirlenmiş olan eğitim kümesindeki elemanlardan yararlanılır. Sınıfı belirlenmek istenen bir eleman her bir sınıftaki diğer tüm elemanlarla karşılaştırılır. Bu elemana en yakın k tanesi seçilir. Seçilen elemanlar en çok hangi sınıfa ait ise sınıflandırmak istediğimiz eleman da o sınıfa aittir. Uzaklığın hesaplanmasında genelde öklid uzaklık veya kosinüs benzerliği formülleri kullanılır [7].

Yapılan çalışmaya ait süreçler Şekil.1'de özetlenmiştir.



Şekil. 1.Süreç Blok Diyagramı

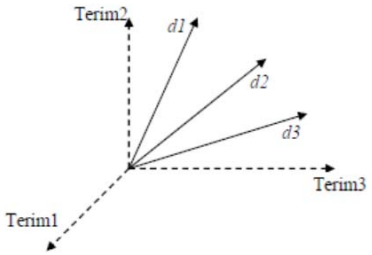
B. Veri Seti

Veri seti için akışkanlar dinamiği, bilimsel endeksler ve tıbbi konulardaki 3 farklı türe ait toplamda 3891 adet makale içeren, literatürde "classic3" veri seti olarak adlandırılan veri kümesi kullanılmıştır.

C. Yöntem

Bir metin dokümanını makine öğrenmesi metotları kullanarak sınıflandırmak veya kümelemek için öncelikle dokümanın hazırlık işleminden geçmesi gerekir. İşleme hazırlama, dokümanı makine öğrenmesi teknikleri için uygun bir duruma getirme sürecidir[8]. Hazırlama işlemi için terim sayma modeli (term count model) veya vektör uzayı-modeli (vector-space model) kullanılır[9].

Bu uygulamada vektör uzayı modeli kullanılmış olup bu modelde, her doküman vektör d ile gösterilmiştir. Vektör d'deki her boyut dokümanların terim uzayında farklı bir terimini belirtir.



Şekil. 2.Dokümanların vektörel gösterimi

Dokümanların hazırlanması 5 farklı işlem ile gerçekleştirilir:

- Dokümanların Ayrıştırılması
- Gereksiz Kelimelerin Temizlenmesi
- Kelime Köklerinin Tespiti
- Terim Ağırlıkları
- Boyutsal İndirgeme

1) Veri Temizleme

Zamirler, edatlar ve bağlaçlar metinlerin birbirleriyle karşılaştırılmasında ayrıştırıcı özelliğe sahip olmadıklarından bunların belirlenip temizlenmesi gerekmektedir. Ayrıca tüm kelimeler küçük harf olarak tek satırlar haline dönüştürülmüş, noktalama ve rakamsal veriler ilgili dokümanlardan çıkartılmıştır. Bu işlemler için "stop-words removal" [10] algoritması kullanılmıştır.

2) Kelime Köklerinin Tespiti

Metinler içinde geçen kelimeler aynı anlam içermesine rağmen cümle içerisinde farklı ekler alabildiklerinden aynı kökten gelen kelimelerin tespit edilmesi gerekmektedir. Bu işlem için "Porter-Stemming" [11] algoritması kullanılmıştır. Bu işlem sonunda kökleri elde edilmiş olan kelimeler öznitelikler kümesi olarak ifade edilecektir.

3) Özniteliklerden Doküman-Terim Matrisinin Oluşturulması

Bu aşamada tüm dokümanların isimleri satırlarda, tüm kelimeler de sütunlarda olacak şekilde denklem (1) 'de ifade edildiği gibi bir matris oluşturulmuştur. Tüm kelimelerin hangi

dokümanda kaç defa geçtiği hesaplanıp bu matriste gösterilmiş ve son satırda da her bir kelimenin tüm doküman kümesinde toplam kaç defa geçtiği hesaplanmıştır.

$$d = (w_1, w_2, \ldots, w_{|N|}) \quad (1)$$

w : i.nci öznitelik (terim),
N : Toplam öznitelik (terim) sayısı.

Bu aşama sonunda 3892 satır ve yaklaşık 29000 sütundan (öznitelik) oluşan bir matris elde edilmiştir [12]. Aşağıdaki tablo 2'de bunun kısa bir örnek gösterimi verilmiştir.

TABLO I. DOKÜMAN-TERİM MATRİS ÖRNEĞİ

	edit	dewey	decim	classif	…
cisi.000001	4	3	2	1	…
cisi.000002	1	0	0	0	…
cisi.000003	0	0	1	0	…
cisi.000004	0	0	0	0	…
...	…	…	…	…	…
Toplam	55	17	26	228	…

4) Benzerliklerin Bulunması ve Başarım Oranının Hesaplanması

Bu aşamada ayrı ayrı vektörel olarak ifade edilen dokümanlar (satırlar), kümeleme işlemi ile eğitim ve test kümesi şeklinde gruplanmıştır. Tüm doküman kümesinin yüzde yirmilik bölümü eğitim kümesi, geri kalanlar test kümesi olacak şekilde seçilmiştir. Eğitim kümesinin her elemanı test kümesinin her elemanı ile hesaplamaya girerek benzerlik oranı elde edilmiştir. K-En yakın komşu algoritmasına göre benzerlik oranının hesaplanması için denklem (2) ve (3) 'deki "Kosinüs Benzerliği" formüllerinden yararlanılmıştır [13].

$$cos\,\emptyset = \frac{d1.d2}{||d1||.||d2||} \quad (2)$$

$$\emptyset = \arccos\left((A \,.\, B) \,/\, (||A||\ ||B||)\right) \quad (3)$$

Kümeleme ve sınıflama algoritmaları kullanmak için iki doküman arasında benzerlik oranının belirlenmesinde bu formül kullanılır. $\emptyset$ değerine göre benzerlik oranı hesaplanır. $\emptyset$ değerinin sıfır çıkması dokümanlar arasında herhangi bir ilişkinin olmadığı, 1'e yakın çıkmasında ise en fazla benzerlik oranına sahip olduğu anlamına gelir.

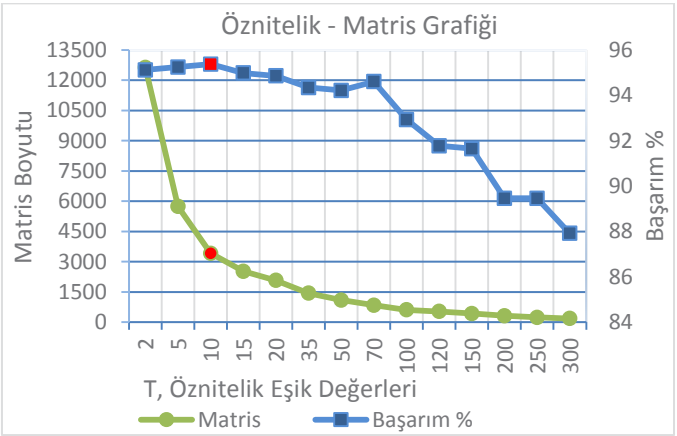
TABLO II. BENZERLİK VE BAŞARIM ORANLARI ÖRNEĞİ

cisi.000001.txt dosyasına en benzer dosya : cisi.001205.txt
Benzerlik oranı : 0.269
cisi.000002.txt dosyasına en benzer dosya : cisi.000916.txt
Benzerlik oranı : 0.513
cisi.000003.txt dosyasına en benzer dosya : cisi.001316.txt
Benzerlik oranı : 0.321
cisi.000004.txt dosyasına en benzer dosya : cisi.001401.txt
Benzerlik oranı : 0.993
……..
Doğru sınıflandırılan doküman sayısı : 742
Yanlış sınıflandırılan doküman sayısı : 36
Başarım Oranı (%) : 95.373

5) Boyut Azaltma

Boyut azaltma başarımı etkileyen önemli bir adımdır. Boyut azaltma sayesinde en ideal işlem sürelerinde başarımı artıran sonuçlar elde edebilecek alt kümeler oluşturulur. Çünkü bu aşamada tüm dokümanlar içerisinde ayırdedici özelliği bulunmayan kelimeler tespit edilerek ilgili matristen çıkarılır ve sütun sayısının boyutu azaltılır.

Yapılan uygulamada aşağıdaki Şekil. 3'de görüldüğü gibi tüm doküman kümesinde toplamda geçen kelime sayısı T ile ifade edilecek şekilde eşik değerleri belirlenmiştir. Bu eşik değerinin altında kalan kelimeler matristen çıkarılarak matrisin boyutunda azaltmaya gidilmiştir. Çıkarılan kelimelerin sınıflandırma işlemi için ayırdedici bir özelliği bulunmamaktadır. Uygulamada 10 eşik değerinde en yüksek başarıma ulaşıldığı görülmüş, boyut daha da azaltıldığında başarım da azalmıştır.

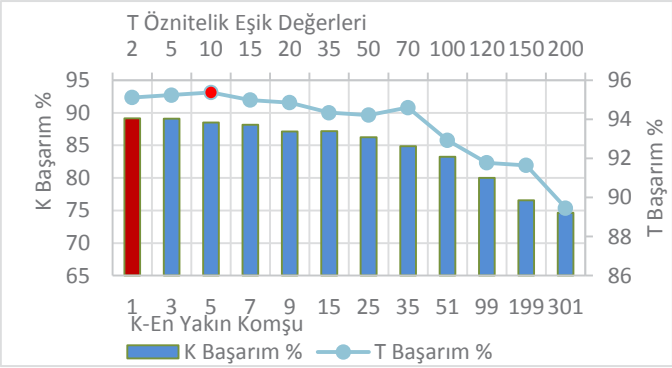


Şekil. 3.Boyut azaltmanın başarıma etkisi

Bu aşama sonunda 3891 satır (doküman) ve yaklaşık 3400 sütun (kelime) içeren boyutu indirgenmiş yeni bir matris elde edilmiştir.

Ayrıca k-en yakın komşu algoritması Şekil. 4'te görüldüğü gibi farklı k değerleri için de uygulanmış, bu veri seti için en yüksek başarım oranı yüzde 95,373 olarak k=1 değerinde elde edilmiştir. Bu veri kümesinde k değeri arttıkça sınıflandırma

başarım oranı azaldığından k=1 ve 10-kat çapraz geçerleme yöntemi ile en yüksek başarıma ulaşılmıştır.



Şekil. 4.Farklı k değerlerinin başarıma etkisi

6) *Ses Dosyalarının Elde Edilmesi*

Ses dosyalarının elde edilmesi için "NetBeans" java derleyici aracına "FreeTTS" kütüphaneleri eklenmiş ve "MBROLA" ses veri tabanı tanımlanmıştır [14].

Başlangıçta 3 türe ait aşağıdaki ses biçimleri Tablo 3'teki gibi tanımlanmıştır. Bu değerler sınıflandırma aşaması sonucunda elde edilen dokümanlarda tanımlanarak ses dosyaları "wav" ses biçiminde elde edilmiştir.

TABLO III. CISI TÜRÜNDEKİ DOĞRU SINIFLANDIRILAN DOKÜMANLAR İÇİN;

Ses Dosyası	:	mbrola_us1 - 16kHz
Cinsiyet	:	Bayan Sesi
Okuma Hız Değeri	:	1.0f
Perde Frekans Değeri	:	180f
Perde Frekans Aralığı	:	22.0f
Ses Yoğunluğu	:	1.0f

Yanlış sınıflandırılan dokümanlar için ses özellik değerlerinde doğru ile yanlış sınıflandırmaları ses dosyalarından ayırt edecek şekilde tablo 4'teki değişiklikler yapılmıştır.

TABLO IV. CISI TÜRÜNDEKİ YANLIŞ SINIFLANDIRILAN DOKÜMANLAR İÇİN;

Ses Dosyası	:	mbrola_us1 - 16kHz
Cinsiyet	:	Bayan Sesi
Okuma Hız Değeri	:	0.5f
Perde Frekans Değeri	:	20f
Perde Frekans Aralığı	:	22.0f
Ses Yoğunluğu	:	0.78f

III. SONUÇ VE ÖNERİLER

Bu çalışmada metinden ses sentezleme ve bunun temel süreçlerinden biri olan metin önişleme aşaması, metinlerin sınıflandırılması ve bu sayede metin türleri hakkında detaylı bilgiler elde edilerek hangi metin türlerinin ne şekilde seslendirileceği konularında önemli sonuçlara ulaşılmıştır. Kullanılan veri kümesinde öznitelik değerleri azaldıkça başarım oranının da azaldığı, anlamsız verilerin temizlenmesi ve boyut indirgeme ile hem işlem sürelerinin kısaldığı hem de başarımın arttığı gözlenmiştir. Bu işlemler sonunda doğru sınıflandırılan dosyaların ses dosyaları belirlenen tonlarda düzgün bir biçimde elde edilirken, yanlış sınıflandırılan dosyaların ses tonlamalarındaki farklılıklar açık bir şekilde ayırt edilebilecek ses biçimleri ile elde edilmiştir.

Gelecek çalışmalarımızda doğallığın artırılarak yeni ses biçimleri elde edebilmek için farklı veri kümeleri ve bunlara bağlı değişen öznitelikler, kullanılacak diğer makine öğrenmesi algoritmaları ile en fazla başarım oranını elde edebilmek için deneysel araştırmalar yapılacaktır.

KAYNAKLAR

[1] Sel, İ., "*Türkçe Metinler İçin Hece Tabanlı Metinden Konuşma Sesntezleme Sistemi Yüksek Lisans Tezi",* Fırat Üniversitesi, Elazığ, 2013.

[2] Güldalı, K., "Türkçe Metin Seslendime Yüksek Lisans Tezi", İstanbul Teknik Üniversitesi, İstanbul, 2009

[3] Eker, B., "*Turkish Text to Speech System, Yüksek Lisans Tezi",* Bilkent Üniversitesi Mühendislik ve Fen Bilimleri Enstitüsü, Ankara, 2002.

[4] Ş. M. Canal, S. Kurnaz, A. E. Yılmaz, "Türkçe Metinden Konuşma Sentezlemede Yaşanan Sıkıntılar ve Çözüm Yöntemleri", Havacılık ve Uzay Teknolojileri Dergisi, cilt 4, sayı 3, s. 47-55, 2010.

[5] Yılmaz, A. E., *"Türkçe Metinden Konuşma Sentezleme Uygulamaları için bir Veri Sözlük Seti ve Yazılım Çerçevesi Önerisi"*, IEEE 17. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, Antalya, 2009.

[6] Basa, G., "*Güngör Basa'nın Web Günlüğü",* http://gungorbasa.blogspot.com.tr/2011/02/speech-synthesis-algorithmskonusma.html adresinden alınmıştır, 2011.

[7] Özkan, Y., "Veri Madenciliği Yöntemleri", Papatya Yayıncılık Eğitim, İstanbul, 2008.

[8] Özgür A., *"Supervised and unsupervised machine learning techniques for text document categorization Yüksek Lisans Tezi"*, Boğaziçi Üniversitesi, İstanbul, 2002.

[9] Salton, G., *"Automatic Information Organization and Retrieval"*, McGraw Hill Text, ISBN:0070544859, 1968

[10] Ronald, *"Remove Stop Words from a File",* https://javaextreme.wordpress.com, 2014.

[11] Porter, M., *"The English Porter stemming algorithm",* https://snowballstem.org, 2015.

[12] Dasan, S., *"Program code for building TDM",* https://snowballstem.org, 2010.

[13] Han, J., Kamber, M., Pei, J., "Data mining: concepts and techniques: concepts and techniques", Elsevier, 2011.

[14] Walker, W., Lamer, P. ve Kwok, P., *"FreeTTS 1.2 - A speech synthesizer written entirely in the Java programming language",* http://freetts.sourceforge.net, 2005.