İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR BİLİMİ VE MÜHENDİSLİĞİ

607 - ÖRÜNTÜ TANIMA PROJESİ I.HAFTA

SESTEN YAZIYA ÇEVRİM (SPEECH TO TEXT, S2T)

DERSİ VEREN: YRD. DOÇ. DR. YAHYA ŞİRİN

MERT YILMAZ ÇAKIR

616816007

Nisan, 2017

Halkalı / İstanbul

İÇİNDEKİLER

A. Çevresel İnceleme

Konuşma tanıma, gelişen teknolojiyle bilgisayarlar ile iletişimimizde, insan sesini aracılar olmadan tanıyarak, bilgisayarlar tarafından okunabilecek bir biçime çevirir. Konuşma tanıma, insanlara konuşma ile cihazları yönetme imkânı tanır. Konuşma tanıma alanında yapılan önceki çalışmalarda, belirli kişi veya grubun sesi ile eğitilip, bu kişi veya gruba ait konuşmaların tanınmasına odaklanılmıştır. Böyle bir konuşmacı bağımlı sistem, yeni bir kişinin sesini, sisteme tanımlamadan çıkarımda bulunamamaktadır. Fakat günümüzde akıllı cihazlarda görülen konuşma tanıma sistemlerinin önemli bir özelliği konuşmacı bağımsız olarak tasarlanması ve dünya genelinde en çok konuşulan dilleri tanımasıdır.

Bu projede, kullanıcı tarafından kelime ya da kelime grubu ile etiketlenen konuşma dizinleriyle kendini dil bağımsız olarak geliştiren bir sistem üzerinde çalışılmıştır. Etiketlenen her bir konuşma ile, dil araştırmaları için yenilikçi bir bakış açısı sayılabilecek bir korpus tabanlı bir konuşma tanıma sistemi oluşturulması önerilmiştir. Bu proje, doğal dil işleme, sinyal işleme, örüntü tanıma gibi bilgisayar bilimleri ile dilbilim, haberleşme, bilgi teorisi gibi alanların kesişiminde yer almaktadır. Proje kapsamında konuşmanın yazıya çevrimi esnasındaki aşamalar incelenecek ve literatürdeki verimli teknikler ile konuşmacı ve dil bağımsız bir sistem önerimi yapılacaktır. Bu projedeki nihai hedef, insanların akıllı cihazlarla doğal ve etkili iletişim kurmalarını sağlamaktır.

Bu bağlamda önce çevresel inceleme yapılmıştır. Birinci bölümde konuşma tanıma modeline giriş yapılmıştır. Çalışmanın ikinci bölümünde konuşma tanıma ile ilgili önceden yapılmış çalışmalara yer verilmiştir. Bu çalışmalar kıyaslanmıştır. Kıyaslanan bu çalışmalardan verimli olan teknikler önerilen projede kullanılacaktır. Üçüncü bölümde önerilen sistem aşamalarıyla incelenmiştir.

I. Giriş

Teknoloji, insan hayatlarını daha basit ve kolay hale getirmektedir. İnsanın, eylemlerini en iyi şekilde ses ile anlatması, insan-makine iletişimini konuşma tanıma sistemlerine yoğunlaştırmaktadır. Konuşma tanıma sisteminin ilk aşamasında, kullanıcının belirli kurallar ile oluşturulan ve kurallarının bilgisayar tarafından bilindiği, birtakım sesli ifadeler alınır. İkinci aşamada bu sesli ifadeler, bilgisayar tarafından anlaşılabilecek formata dönüştürülür. Konuşma tanıma; akıllı cihazlarda sesli komut uygulamaları, akıllı ev sistemleri, sesli komutlar ile sağlanan güvenlik sistemleri, eğitim sistemleri, Etkileşimli Ses Yanıtı (Interactive Voice Responce) ve Sesli Yanıt Ünitesi (Voice Response Unit) gibi birçok alanda gelişmeye ve geliştirilmeye devam etmektedir.

Konuşma tanıma sistemleri üzerine yapılan önceki uygulamalar, konuşmacı bağımlı olarak eğitilip, eğitilen konuşmalara göre kişi tanıma üzerine olmuştur. Bu tip konuşmacı bağımlı sistemlerde tanıtılmayan kişiler için sistemin eğitilmesi gerekmektedir. Güvenlik, yetkilendirme gibi alanlarda tercih edilen bu sistemler, imkânlar ve zamanın sistemin konuşmacı bağımlı eğitilmesine göre yeterli olduğu uygulamalar için kullanılır. Konuşmacı bağımsız sistemler ise konuşmacı bağımlı sistemlere göre karmaşıklığı fazla ve daha zor oluşturulan sistemlerdir. Fakat, bu sistemler, konuşmacı bağımlı sistemler gibi bir şablon güncellemesine ihtiyaç duymadan herhangi birinin sesini tanımaya olanak sağlar.

Konuşmacı bağımsız sistemler, kaydedilen çok sayıda ses örnekleriyle ön öğrenmeden geçirilerek kullanılır. Böyle bir sistemin dezavantajı ise herhangi bir dil için tüm konuşmacı varyasyonlarını modellemenin zor olmasıdır. Akıllı telefonlar da dahil olmak üzere birçok alanda örneği bulunan konuşma tanıma sistemlerinin önemli özelliklerinden bir tanesi konuşmacı bağımsız olması ve günümüzde en çok konuşulan dilleri tanımasıdır. Bu çalışmada da konuşmacı ve dil bağımsız bir konuşma tanıma sistemi için literatür taraması yapılmış ve verimli teknikler ile konuşma tanıma önerimi yapılmıştır. Önerimi yapılan sistemin uygulaması yapılacaktır.

II. Önceki Çalışmalar

2016 yılında Karthikeyan V. ve arkadaşları, ses uygulamaları için konuşma tanımanın performans karşılaştırması üzerine çalışmışlardır. Çalışmalarında özellikle görme zorluğu çeken kişiler için cihaza dokunmadan tüm cep telefonu uygulamalarını kullanabilmeleri için konuşmacı bağımsız sistem üzerinde durmuşlardır. Önerilen sistemde konuşma özellikleri, MFCC (Mel Frequency Cepstral Cofficients) tekniği kullanılarak çıkartılmıştır. DTW (Dynamic Time Warping) ve HMM (Hidden Markov Model) / VQ (Vector Quantization) kullanılarak şablon üretimi gibi iki farklı sınıflandırma modellemesi yoluyla değerlendirilmiştir. MFCC özellikleri ile HMM / VQ sınıflandırma modeli, sesli algılamalar için diğer metotlara göre daha yüksek olarak % 82.77 tanıma oranı vermiştir [1].

2016 yılında Imtiaz M.A. ve Raja G., otomatik konuşma tanıma (ASR) sistemi akustik konuşma sinyallerini kelimelerin dizisine dönüştürmek olarak tanımlayarak, MFCC, DTW ve K-En Yakın Komşu (KNN) teknikleri kullanılarak izole sözcük yapısına dayanan ASR sisteminin bir yaklaşımını sunmuşlardır. Konuşma sinyallerinin belirgin özelliklerini yakalamak için kullanılan Mel-Frekans ölçeği ile konuşma özellikleri MFCC kullanılarak çıkartılmıştır. DTW, konuşma özelliği eşlemesi için uygulanmıştır. KNN sınıflandırıcı olarak kullanılmıştır. Deney düzeneğinde, beş konuşmacıdan toplanan İngilizce kelimeler bulunmaktadır. Bu kelimeler, akustik olarak dengeli, gürültülü olmayan bir ortamda söylenmiştir. Önerilen ASR sisteminin deneysel sonuçları, karışıklık matrisi adı verilen matris formunda elde edilmiştir. Bu araştırmada elde edilen tanıma doğruluğu % 98.4 olmuştur [2].

2016 yılında Bakır, Almanca ses biçim ve özelliklerine bakılarak konuşmacının cinsiyetinin otomatik olarak tanınması için bir sistem tasarlamıştır. 50 erkek ve 50 kadından Almanca farklı uzunlukta kelime ve cümle ile yaklaşık 3000’ e yakın ses örneği alınmıştır. Ses örnekleri üzerinde özellik vektörleri, MFCC kullanılarak elde edilmiştir. Elde edilen ses örnekleri HMM, DTW ve GMM (Gaussian Mixture Model) yöntemleri ile eğitilmiştir. Test aşamasında ise ses örneklerine bakılarak verilen ses örneğinin cinsiyeti belirlenmeye çalışılmıştır. GMM ile %84.26 oranında, DTW ile %87.37 oranında başarım gerçekleşirken, HMM ile %98.34 oranında başarım sağlanmıştır [3].

Önerilen proje için literatür araştırması yapılmıştır. Araştırmalar, yapılması hedeflenen projede kullanılacak olan tekniklerin tespitinde önemli rol oynayacaktır. Son yıllarda yapılan ve yukarıda anlatılan üç akademik çalışma detaylı incelenmiştir. İncelemeler neticesinde hedeflendiği gibi izole ve konuşmacı bağımsız sistemlerde özellik çıkarımında MFCC tekniği başarılı olarak kullanılmıştır. Ayrıca elde edilen özelliklerin sınıflandırılmasında yukarıda da başarım oranı verilen HMM tekniğinin verimli olduğu gözlemlenmiştir. Bu sebeple bu tekniklerin projede kullanılması hedeflenmiştir.

III. Önerilen Sistem

Konuşma, insanlar arasında hızlı, etkin ve çok yönlü bir iletişim aracıdır. Konuşma içerisindeki bilgiler, karmaşık bir biçimde kodlanmıştır ve insanlar tarafından şifresi çözülebilmektedir. Bu insan kabiliyeti, araştırmacılara bu yeteneği taklit edecek sistemleri geliştirmeye ilham kaynağı olmuştur. Ses bilgisi uzmanlarından mühendislere kadar birçok araştırmacı, konuşma sinyalindeki bilgileri çözmek için çeşitli alanlarda çalışmaktadırlar. Bu alanlara, konuşulanların sese göre belirlenmesi, konuşulan dilin keşfedilmesi, konuşmanın aktarılması, konuşmanın tercümesi ve konuşmanın tanınması örnek olarak verilebilir.

Konuşma tanıma, bir kişinin bir mikrofona veya benzer bir donanıma ne söylediğini tanımlama ve anlamını metin, resim veya herhangi bir olay gibi gerekli herhangi bir biçimde yansıtan bir süreçtir. Konuşma tanıma, birçok araştırmacının uzun yıllardır üzerinde çalıştığı bir alandır. Bu alanda, konuşmacı dil bilgisi mesajı ile ilgilenmektedir. Konuşma tanıma, bir metnin dikte edilmesinden gerçek zamanlı olarak bir televizyon yayını için altyazı üretmeye kadar birçok uygulamayı içerir. Konuşmacı tarafından dilsel, fizyolojik ve çevresel birtakım faktörlere bağlı olarak konuşmada değişkenlikler gözlenebilir. Böylece araştırmacılar, bir insan yeteneği olmasına karşın konuşmadan bilgi çıkarmanın basit bir süreç olmadığını tecrübe etmişlerdir. Bu tecrübeler ile araştırmacılar, konuşma sinyalinden ilgili bilgileri güvenilir bir şekilde çıkartmaya çalışmaktadırlar.

Günümüzde konuşma tanımanın piyasada çeşitli etkileşimli uygulamalarına ihtiyaç artmaktadır. Konuşma tanıma sistemleri ile uygulandığı alana göre, kullanım kolaylığı, veri toplama hızı, hareket serbestliği ve uzaktan veri giriş imkanı sağlanabilir. Konuşma tanıma literatürde geniş yer tutan bir örüntü tanıma problemidir ve bu problem özelinde elde edilen sonuçlar tüm literatüre katkı sağlar niteliktedir. İnsan iletişiminin en doğal biçimi olan konuşma yönteminin akıllı cihazlara yönelmesiyle, sesli komutlar ile özellikle ellerin ve gözlerin meşgul olduğu otomobillerde kolaylık sağlanabilir. Konuşma tanıma, çeşitli otomasyon ve güvenlik sistemleri, akıllı cihazların (telefon, tablet, vs) sesli komut ile kontrolü, bilgisayar tabanlı telesekreter gibi birçok kullanım alanı olabilecek bir uygulamadır.

Konuşmanın yazıya çevrilmesi için sesli ifadelerin, öncelikle bu sürece hazırlanarak bilgisayar destekli olarak tanıma sürecine dahil edilmeleri gerekmektedir. Bu amaçla sesli ifadelerin bir mikrofon aracılığıyla örneksel sinyallere dönüştürülmesi, sayısal olarak işlenen bu sinyallerin gerekirse filtrelenmesi, etiketlenmesi (örneğin sesler, fonemler, kelime ya da kelime grubu olarak) ve tanıma işlemlerine taban oluşturacak sınıflandırma teknikleriyle parametrik yapılar ya da yalın modellerle ifade edilen biçimlere dönüştürülmesi gerekmektedir. Önerilen projede, kullanıcı bağımsız bir sistem üzerine çalışılmıştır. Böyle bir sistemin zorluklarını göz önüne alırsak, bir ifade üzerine birden fazla kişinin ses kaydı alınması, sistemin verimini arttıracaktır. Önerilen projede, sistem kendini öğrendiği her yeni bilgiyle geliştirmelidir. Bu özellik ile sistem hangi dil üzerinden eğitilmeye başlarsa, bu dil üzerinden sistem gelişecektir. Konuşma tanıma sistemi, algılanan sesin yazıya çevrilmesinde, soyutlandığı zaman iki aşamada incelenir. Bunlar sistemin tanıma evresi ve eğitim evrelesidir.

Konuşma tanıma, büyük potansiyellere sahiptir ve yetersizlikleriyle insanlar için eğitimsel uyum süreci vardır. Çoğu kez bir konuşma tanıma sistemi çalışmazsa, bu kullanıcının davranışına ve bilgisine bağlıdır. Bunun gibi yetersiz bilgiyi önlemek için iki ölçüm vardır. Kullanıcılar sistemi, ses ile çalıştırmak için iyi hazırlanmış olmalıdır ve güncel teknolojileri bilmelidirler. Sistemin

eğitiminde verilen girdiye göre şekillenen çıktı bilgisi kıyaslanarak hesaplama yapılır. Hata oranın eğitim ile küçülmesi gerekmektedir. Burada eğitim evresi sınıflandırma fazında yapılır. Belirlenen teknik ile sistem eğitimine ve buna bağlı olarak da hata oranın azalmasına devam eder.

Algılama cihazları ile saptanan ya da belirli bir formattaki konuşma, tanıma evresinde işlem görür. Bu aşamada kullanıcıdan alınan konuşma, tanıma evresindeki aşamalar ile yazıya çevrilir. Tanıma evresi, konuşmanın algılanması, sayısal olarak anlamlı özellik vektörlerine çevrilmesi ve sınıflandırma ile yazıya çevrilmesi işlemidir. Bu projede konuşma tanıma süreci, konuşmacı tarafından söylenen kelimelere karşılık gelen özellik vektör dizilerinin çıkartılması ile başlanır. Konuşma tanıma sistemlerinde özellik vektörlerinin elde edilmesi sırasında literatürde kullanımı ve performansı yüksek olan MFCC algoritması önerilmiştir.

Özellik vektörlerinin bulunmasından sonra, bu kelimelere karşılık gelen istatistiksel model ile veritabanı oluşturulur. Söylenen kelimeye karşılık, tüm veritabanı içerisinden arama yapılır ve verilen sinyale en uygun eşleşme seçilir. Her yeni eklenen ve sisteme kaydedilen ses dosyaları yazısıyla etiketlenerek VQ ile sistemin kod kitapları oluşturulur ve vektörler arası uzaklıklar hesaplanır. Önerilen projede, kullanıcı ses kaydıyla konuşma metnini etiketleyerek sistemin kod kitabına kaydedilmesini sağlar. Kaydedilen metin sınıflandırma algoritmaları ile güncellenir ve sistem yeni olasılıksal modeller üretir. Üretilen bu modeller daha sonra sistemin test edilmesi aşamasında en yüksek olasılıklı eşleşmeyi yaparak konuşmaya en uygun metni ekranda gösterir.

HMM’ nin buradaki amacı gözlemlenen durumlara karşı olan durumları tahmin etmektir. HMM’de durumlar, gözlemler ve durumlar arası geçişler vardır. Konuşma tanımada gözlemler, ses sinyalinden elde edilen özellik vektörlerinden oluşur. Durumlar ise temel olarak alınan ses ifadelerinin karşılığı olan birimlere denk gelir. Bu durumda konuşma tanımada ki amaç saklı olan durum dizisini gözlemlerden yararlanılarak çıkarılan olasılıklardan bulmaya çalışmaktır. Her ses ifadesi için ayrı bir model tanımı yapılır. Her ses ifadesi bir model olarak düşünülürse gerçek zamanlı olarak gelen ses ifadeleri bu modellerin art arda sıralanması ile modellenir. Bu durumda her bir ses ifadesinin son durumda bir sonraki ses ifadesinin ilk durumuna bir geçişi söz konusudur. Böylece konuşma tanıma HMM ile gerçekleşmiş olur.

Burada sistemin her bir söz dizimi için eğitilmesi, HMM Baum Welch algoritması ile durumlar arasında olasılıkların hesabıdır. VQ ile bu durumsal olasılıklar için istatistiksel bir çerçeve sunulmaktadır. Bu yaklaşım ile HMM durumları üzerinde VQ Kod Defterinin optimum dağılımı gerçekleştirilir. Bu önerilen sistemde HMM’nin dağıtılmış VQ’sudur. Literatür araştırmalarımız ile konuşma tanıma üzerine önerdiğimiz çalışmada, konuşmacı bağımsızlık, dil bağımsızlık ve sistemsel verimliliğin/doğruluğun arttırılması hedeflenmiştir. Önerdiğimiz konuşma tanıma sisteminde sınıflandırma aşamasında, sinyal modellerin, özünü bilmeden sinyalin kaynağıyla ilgili modelleme yapabilmesinden HMM önerilmiştir. HMM'in geri görünümünde çalışan saklı bir Markov işlemi bulunur. Bu modeller gözlem vektörleri üretir ve HMM için gözlem dizileri oluşur.

Viterbi algoritması, özellikle Markov bilgi kaynakları ve HMM bağlamında, gözlemlenen olayların bir dizilimiyle sonuçlanan, gizli durumların en olası sırasını bulmak için uygundur. Dinamik programlama algoritması ile bulunan en uygun şablon model yazı olarak ekrana yazılır. Her söyleniş ideal halde bir HMM' ye sahip olmalıdır. Fakat kimi zaman bu olmaz, bu sebeple sözcük bazında HMM' lerimiz olmalıdır. Konuşmacı tarafından söylenen kelimeyle karşılaştırıp eşleştirmek için bir HMM, veritabanında bulunan bütün kelimeler için en iyisini gerçekleştirmelidir.

KAYNAKLAR

1.http://www.ajetr.org/vol16/no1/n06.pdf

2.http://ieeexplore.ieee.org/abstract/document/7878163/

3.https://www.google.com.tr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjB0dvWvqLTAhWIxRQKHc1dC1IQFggiMAA&url=http%3A%2F%2Fdergipark.ulakbim.gov.tr%2Fapjes%2Farticle%2Fdownload%2F5000184477%2F5000170530&usg=AFQjCNGCBobGf_DyqjHfvi6lu2gxh7RBWg&sig2=WzgxWarmgOxAJTYw2wy0aA&bvm=bv.152180690,bs.1,d.bGs