



T.C.
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**Parçacık Sürü Optimizasyonu Kullanarak
Makine Öğrenmesi İle Konuşma Tanıma
Performansının Artırılması**

Arzo MOHAMMED MAHMOOD MAHMOOD

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Temmuz-2019
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Arzo MOHAMMED MAHMOOD MAHMOOD tarafından hazırlanan “Parçacık Sürü Optimizasyonu Kullanarak Makine Öğrenmesi İle Konuşma Tanıma Performansının Artırılması” adlı tez çalışması 22/07/2019 tarihinde aşağıdaki jüri tarafından oy birliği ile Selçuk Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

Başkan

Doç. Dr. Barış KOÇER

Danışman

Dr. Öğr. Üyesi Ersin KAYA

Üye

Dr. Öğr. Üyesi Ayşe Merve ACILAR

İmza



Yukarıdaki sonucu onaylarım.

Prof. Dr. Mustafa YILMAZ
FBE Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.



İmza

Arzo MOHAMMED MOHAMMED MAHMOOD
Tarih: 22/07/2019

ÖZET

YÜKSEK LİSANS TEZİ

Parçacık Sürü Optimizasyonu Kullanarak Makine Öğrenmesi İle Konuşma Tanıma Performansının Artırılması

Arzo MOHAMMED MAHMOOD MAHMOOD

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: DR.ÖĞR.ÜYESİ.Ersin KAYA

2019, 61 Sayfa

Jüri

Dr.Öğr.Üyesi ERSİN KAYA
Dr.Öğr.Üyesi BARIŞ KOÇER
Dr.Öğr.Üyesi AYŞE MERVE ACILAR

Konuşma tanıma, insan sesinin bilgisayar tarafından algılanması olarak tanımlanmıştır. Konuşma tanıma öznel bir olgudur. Bu alanda birçok araştırmalar yapılmış olmasına rağmen hala birçok sorunla karşılaşılmaktadır. Bu alanda çeşitli ilerlemeler sağlanmıştır ve her amaç için farklı teknikler kullanılmaktadır. Bu tezde, Parçacık Sürüsü Optimizasyonu (PSO) ile k-en Yakın Komşuluk Algoritması (KNN), Destek vektör makinaları (DVM) ve yapay sinir ağları (YSA) sınıflandırma tekniklerini bir arada kullanılmıştır. Doğrusal Öngörüm Kodlaması (LPC), konuşma sinyali özelliklerinin çıkarılması için kullanılmaktadır ve destek vektör makinelarının (SVM) sınıflandırmada önemli bir noktası olan öğrenme aşamasında sezgisel bir algoritma olan Parçacık Sürü Optimizasyonu (PSO) kullanılarak sınıflandırmanın başarısı arttırılmıştır.

Bu tezde farklı yaştaki farklı kişilerin sesleri sessiz ve gürültüsüz bir ortamda kaliteli bir mikrofon ile kaydedilmektedir. Her biri 5 kelimededen (Back, go, left, right ve stop) oluşan ve içerisinde 12 kişinin bulunduğu bir veri seti kullanılmaktadır. Eğitim seti 60 örnekten test seti ise 40 örnekten oluşmaktadır ve kelime süresi 1 saniyedir. PSO ile optimize edilmiş konuşma tanıma sisteminde, SVM, KNN VE YSA olmak üzere ayrı ayrı üç sınıflandırma kullanılmıştır .

Anahtar Kelimeler: Konuşma Tanıma, Lpc, Pso, Ses, Dvm, Ysa

ABSTRACT

MS THESIS

Improved speech recognition performance with machine learning using particle swarm optimization

Arzo MOHAMMED MAHMOOD MAHMOOD

**THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF
SELÇUK UNIVERSITY**

Department of COMPUTER ENGINEERING

Advisor: MEMBER OF DR. Ersin KAYA

2019, 61 Pages

July

DR. ERSİN KAYA

DR. BARIŞ KOÇER

DR. AYŞE MERVE ACILAR

Speech recognition is defined as the recognition of human voice by computer. Speech recognition is a subjective phenomenon, although many studies have been conducted in this field, many problems are still encountered. Various advances have been made in this field and different techniques are used for each purpose. In this study, Particle Swarm Optimization (PSO) and Support Vector Machine (SVM), knearest neighbors algorithm (KNN), and Artificial Neural Network(ANN) techniques are used together. Linear Prediction coding (LPC) is used to extract speech signal characteristics and Support Vector Machines (SVM), an important point in the classification of the learning stage, an intuitive algorithm, Particle Swarm Optimization (PSO) has been increased the success of the classification.

The voices of different people of different ages are recorded with a quality microphone in a quiet environment. A data set of is used, each consisting of 5 words (Back, go, left, right and stop) and 12 people 60instance of training set and 40 instance of test set each word's duration is 1 second spoken by different people. In the PSO-optimized speech recognition system three different classifications,SVM,KNN and ANN were used, and if compared, the traditional classification used was better compared to SVM.

Keywords: Ann, Lpc, Pso, Sound, Speech Recognition, Svm.

ÖNSÖZ

Tez çalışmamda, benden yardımlarını esirgemeyen başta danışmanım Dr. Öğr. Üyesi ERSİN KAYA olmak üzere, maddi ve manevi her zaman yanımda olan aileme ve benden desteklerini hiç esirgemeyen değerli iki arkadaşım Zahide UÇAK ve Abdullah FADHİL'a sonsuz teşekkür ederim.

Arzo MAHMOOD
KONYA-2019



İÇİNDEKİLER

ÖZET	iv
ABSTRACT	v
ÖNSÖZ	vi
İÇİNDEKİLER	vii
SİMGELER VE KISALTMALAR	ix
1. GİRİŞ	1
2. KAYNAK ARAŞTIRMASI	4
3. MATERYAL VE YÖNTEM	7
3.1. Ses Nedir?	7
3.1.1. Ses hangi ortamlarda yayılır?	7
3.1.2. Ses neden boşlukta yayılmaz?	8
3.1.3. Sesin yayılma hızı	8
3.1.4. Konuşma tanıma	8
3.2. Yapay Sinir Ağları (YSA)	9
3.2.1. Yapay sinir ağlarının sınıflandırılması	10
3.2.2. Yapay sinir ağının türleri	10
3.2.3. Yapay sinir ağının avantajları	11
3.3. Doğrusal Öngörü Kodlaması (Linear Prediction Coding (LPC))	11
3.3.1. LPC katsayılarının ölçümü	12
3.3.2. LPC-motivasyon	14
3.3.3. LPC - Birinci otomatik korelasyon	14
3.3.4. LPC Uygulamaları	15
3.4. Parçacık Sürü Optimizasyonu (PSO).....	16
3.4.1. Pbest- Gbest	18
3.4.2. Avantajları	18
3.4.3. Dezavantajları	19
3.4.4. Yöntemin uygulanışı	19
3.4.5. Hız bileşenleri	20
3.4.6. Parçacık sürü optimizasyonu ile geometri arama	21
3.4.7. PSO 'nun tartışılan yönleri	23
3.5. Destek Vektör Makineleri (SVM)	24
3.5.1. Verilerin sınıflandırılması	25
3.5.2. Verileri dönüştürmek	26
3.5.3. Öğrenme	27
3.5.4. Çok sınıflı destek vektör makineleri	27
3.5.5. Destek vektör makinelerin avantajları ve dezavantajları	28
3.5.6. Yumuşak marj sınıflandırıcı	29
3.5.7. Çekirdek numarası	29
3.5.8. Çekirdek İşlevleri	32

3.5.9. Regresyon için SVM.....	32
3.5.10. Destek Vektör Makinelerinin Uygulamaları.....	33
3.6. K-NN Yakın Komşu Algoritması	34
3.6.1. KNN algoritması nasıl çalışır	35
3.6.2. K faktörünü nasıl seçeriz?.....	36
3.6.3. Ağırlıklı en yakın komşu sınıflandırıcı	38
3.7. Konuşma Tanıma Uygulamalarında Sınıflandırma Yöntemleri	38
3.7.1. Konuşulan konuşma türleri	39
3.7.2. Uygulama çerçevesi	40
3.7.3. Performans değeri ölçütleri.....	40
4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA.....	42
4.1. Önerilen Metod	42
4.2. Araştırma Sonuçları	46
5. SONUÇLAR VE ÖNERİLER	49
5.1 Sonuçlar	49
5.2 Öneriler	50
KAYNAKLAR	51
ÖZGEÇMİŞ	54

SİMGELER VE KISALTMALAR

Simgeler

£ : pound Sign

% : percent sign

Hz : Hertz

DB : desibel

Kısaltmalar

YSA :YAPAY SİNİR AĞLARI

ANN:Artificial Neural Network

PSO : Parçacık Sürü Optimizasyonu

LPC: Doğrusal Öngörüm Kodlasmsı

LPC : Linear Predictive Coding

DVM: Destek Vektör Makinaları

SVM : Support Vector Machine

KNN : k-En Yakın Komşuluk Algoritması

KTS : Konuşma Tanıma Sistemleri

CPS :Cycle Per Second

DB : desibel

MATLAB : MATrix LABoratory

1. GİRİŞ

Konuşma, insanların kendilerini ifade etmesi ve birbirleriyle arasında bilgi alışverişi sağlamasının en etkili ve doğal yol olarak tanımlanabilir. İnsanların ve makinelerin arasındaki iletişimini sağlamak ve akıllı bilgisayarlar elde etmek için makinelerin "işitmesi anlaması" ve buna göre davranması önemlidir. Bu durumda, insan ile bilgisayar arasındaki iletişimi sağlamak için bir bilgisayarda "konuşma tanıma" gerekli hale gelmiştir. Ses tanıma sistemleri, insanlar arasındaki sesli iletişim sürecinde dinleyicinin işlevlerini yapay olarak yerine getirmeye çalışır. Güvenilir konuşma tanıma, birçok teknikle ilgili zor bir sorundur ve modern yöntemlerin etkileyici bir doğruluğa ulaşmasına rağmen hala bazı sorunları vardır.

Konuşmacı sayısına bağlı olarak konuşma tanıma sistemleri 2'ye ayrılmıştır. Tek bir kişinin sesiyle eğitilen sistem, kişiye bağımlı sistem olarak adlandırılır (Tunali ve Doğruel, 2005). Genel olarak bu sistemlerin başarı oranı çok yüksektir. Kişiden bağımsız (kişiye bağlı olmayan) sistemlerde bir çok insanın seslerini tanımak mümkündür (Rabiner ve ark., 1979). Bu sistemleri, herkesin çalıştırabileceğinden dolayı, çok fazla sayıdaki insanın sesini sisteme tanıtmak gerekebilir. Bundan dolayı bu durum, konuşma tanımanın başarısını olumsuz yönde etkileyebilir. Konuşma tanıma alanında son yıllarda yapılan çalışmalar hızla artmaktadır. Konuşma tanıma çalışmalarının çoğunda ses birimleri olan fonem tabanlı, hece tabanlı ve kelime tabanlı birimler kullanılmıştır.

Konuşma tanımayı zorlaştıran bazı faktörlere bakıldığında; konuşmacının insandan insana değişmesi, kullanılacak kelimelerin fazla olması, konuşmaya harici seslerin karışması, müzik gürültü gibi ya da iki konuşmanın birbiriyle aynı olmaması, günlük konuşma dili ve aksan nedeniyle kelimelerin söylenişlerinin çeşitlenmesi örnek olarak "geleceğim" sözcüğünün "geleceğim, gelicem, gelcem" şekillerinde söylenmesi gibi etkenlerin olduğu görülür. Konuşma Tanıma Sistemleri (KTS)'nin amacı, bu zorlukları en aza indirmek ve tanımayı kolaylaştırmaya yardımcı olmaktadır. Günümüzde KTS, birçok değişik alanlarda kullanılabilir (Yılmaz, 1999; Yakar, 2016). Genel olarak bakıldığında bir Konuşma Tanıma Sistemi'nin görevi, insanlara ait ses sinyallerini alıp, bu sinyaller üzerinde birkaç işlemler yapmak, yapılan bu işlemlerden sonrada ses sinyallerinin hangi kelimelere karşılık geldiğinin tespit ederek yazıya dökmektir (Aksoylar ve ark., 2009). Bundan dolayı, bir Konuşma Tanıma

Sistemi, girdi olarak bir konuşma sinyalinin alır, çıktı olarak ise bir metin üretmektedir (Nádas ve ark., 1988)

Geçmişten günümüze kadar gelen bütün uzmanlar insan davranışlarını taklit edebilen icatlar geliştirmek istemişlerdir. Bu isteklerin en dikkat çekici olanı ise insanlar gibi konuşabilen robotlar geliştirmek istemeleridir.

Konuşma tanıma teknolojilerinin, önemi her geçen gün daha da artmaktadır. Bu teknoloji sayesinde hayatın hem kolaylaştığını hem de zamandan tasarruf etmenin önemini görmekteyiz.

İlk olarak geliştirilmiş olan konuşma tanıma sistemleri sadece rakamları anlayabilmekteydi. Bu alandaki bilim adamları bu rakamların sayılarına odaklanmışlardı. Bell Laboratuvarlarında 1952 yılında geliştirilen "Audrey" adındaki konuşma tanıma sistemi, rakam bir kez söylendiğinde tanıyabilmekteydi. 1962 yılında yapılan Dünya fuarında IBM şirketinin "Shoebos" adlı konuşma tanıma makinesi, 16 tane İngilizce kelimeyi tanıyabilmekteydi. 1950 ve 1960'larda Sovyetler Birliği, ABD, İngiltere ve Japonya'daki laboratuvarlar da ünlü ve ünsüz harfleri tanıyan donanım tabanlı konuşma tanıma sistemleri geliştirilmiştir.

ABD Savunma Bakanlığı'nın DARPA Konuşma Anlama Araştırma Programı (SUR) 1971'den 1976'ya kadar devam etti ve konuşma tanıma konusunda büyük ilerlemeler kaydetmiştir. 1970'lerde Carnegie Mellon Üniversitesi tarafından geliştirilen "Harpy" adındaki konuşma anlama sistemi 1011 İngilizce kelimeyi anlayabilmekteydi. Bu yıllarda "Threshold Technology" olarak adlandırılan ilk ticari şirket kuruldu ve farklı insanların seslerini anlayabilmek için ilk ticari konuşma tanıma sistemi (VIP-100) piyasaya tanıtıldı. 1980 yıllarına kadar olan konuşma tanıma sistemleri birkaç yüz kelime tanıyorken artık günümüzde bu sayı birkaç bin kelimeye kadar artmıştır.

Bu tezde ses tanıma ile ilgili bu tekniklerden birisi ele alınmış ve örnek bir ses tanıma sistemi geliştirilmiştir. Konuşma sesleri ses kaydı kullanılarak bilgisayara aktarılmış ve bu veri MATLAB'da hazırlanmış program yardımı ile işlenmiştir. Ses tanıma süreci, sesin kaydedilmesi ile başlar, sesin işlenmesi, öz niteliklerinin çıkarılıp kaydedilmesi, karşılaştırma ve eşleştirme yapılarak sesin tanınması ile son bulur. Bu süreç boyunca ses çeşitli aşamalardan geçer. İlk olarak sesin sayısallaştırılması işlevi gerçekleştirilmektedir. Sesin sayısallaştırılmasında sonra sesin kodlanması gerçekleştirilir. Bu tezde farklı yaşta farklı kişilerin sesleri sessiz ve gürültüsüz bir ortamda kaliteli bir mikrofon ile kaydedilmektedir. Her biri 5 kelimedenden oluşan (Back, go, left, right ve stop) ve içerisinde 12 kişinin bulunduğu bir veri seti kullanılmaktadır.

Eđitim seti 60 6rnek ve 12 kiřiden test seti ise 40 6rnek ve 8 kiřiden oluřmaktadırdır, kelime s6resi 1 saniyedir bu kiřiler tarafından konuřulmaktadır. Bu tezde kullanılan 12 kiřiden 7'si kadın 5'i ise erkeklerden oluřmaktadırdır. Deneysel alıřmada kullanılan kiřilerden 3'6 30, 4'6 25, 3'6 20 ve 2'si ise 10 yařındadırdır. LPC ile ıkarılan 6zelliklerinin sayısı 200 d6r. PSO ile her bir ses dosyasının 200 6zelliđinden uygun bir k6me belirlenir ve alıřmada kullanılan SVM, KNN ve ANN sınıflandırıcıları eđitir, daha sonra seilen her 6zellik k6mesinin dođruluđu hesaplanır ve seilen bu 6zelliklerle test setinin sınıflandırılması yapılır.



2. KAYNAK ARAŞTIRMASI

Derin sinir ağıları (DNNs) konuşma tanıma için akustik modellemede büyük başarı elde etmiştir. Bunun birlikte, DNN switch sigmoid nöronlar antrenman sırasında kaybolan gradyan probleminden muzdarip olabilir. Maksant nöronlar sigmoidneurons için umut verici alternatiflerdir. Amazon Neuron'un yerel bir bölgesinde maksimum değer seçilerek, eğitim sürecinde sürekli derecelendirmelerle sonuçlanan aktivasyon gerçekleştirilir. Bu yazıda, maksimum nöronları akustik modelleme için iki popüler DNN yapısı, yani konvolüsyonel nöral ağ (CNN) ve uzun süreli kısa süreli bellek (LSTM) tekrarlayan sinir ağı (RNN) birleştirilmiştir. Modeller için en uygun ağ yapıları ve eğitimleri incelenmiştir. Önerilen Modeller, altı dil koleksiyonunda karşılık gelen CNN veya LSTM RNN taban çizgileri üzerinde% 2.5-6.0 nispi iyileşmeler elde etmiştir. Bu veri setlerinde en son teknoloji ürünü sonuçlar, sistem kombinasyonundan sonra elde edilmektedir (Cai ve Liu, 2016).

Hangi çalışma, güvenlik biyometrik tanıma sisteminde Mel Frekanslı Cepstral Katsayıları ile Yapay Sinir Ağları algoritmalarını uygulamıştır. MATLAB simülasyon ortamında 10 kişi için tanıma doğruluğu oranı% 99 olarak bulunmuştur (Jiang ve ark., 2009).

Tasarlanan Visiual Basic 6 kullanıcı arayüzü Gizli Markov Modeli kullanıcı kitini kullanmıştır. Gürültülü bir ortamda, uzaktan kumandalı bir araba konuşma komutları ile kontrol edildiğinde,% 98 hassasiyet başarısını elde ederken, gürültülü bir odada% 44'e kadar düşmüştür (Leechor ve ark., 2010).

Avigator aracını tasarlamak için Mel Frekanslı Cepstral Katsayıları ve ortak vektör yaklaşımı algoritmalarını uyguladı. Hidden Markov Model ile yaklaşık sonuçların elde edilmesinde yaklaşık% 100 başarı sağlandı (Babu ve ark., 2012).

İzole konuşma tanıma için yapay sinir ağıları kullanılmıştır. Konu, ön işleme bölümünden oluşan iki aşamada incelenmiştir. Dijital Sinyal İşleme (DSP) teknikleri ve Yapay Sinir ile post-processing bölümü Ağlar (YSA). Bu iki bölüm kısaca açıklandı ve konuşmacı tanıyıcıları kullanılmıştır. Matlab'da farklı YSA mimarileri uygulanmıştır. Üç farklı sinir ağı modeller; Çok katmanlı geri yayılım, Elman ve olasılıksal sinir ağıları tasarlandı. İlgili literatürde benzer çalışmalarla yapılan performans karşılaştırmaları bizim için önerilen YSA yapıları tatmin edici sonuçlar vermektedir (Dede ve Sazlı, 2010).

Yapay zeka ile abonenin ses analiz işleminin abonenin SIMKart aldığı operatörün yetkili servisinde yapılabileceğini öne süren Önder Türkoğlu, “Mobil Haberleşmede Ses Tanıma ve Eğitime Etkileri” isimli çalışması ile ses tanıma yönelik bir çalışma sunmuştur (Grieder ve Kinsner, 1994; Benesty ve ark., 2008).

Tunçkanat ve ark. (2003), Yapay Sinir Ağları ile Konuşmacı Tanıma isimli çalışmasında, konuşmacı tanımada YSA'nı kullanarak Levenberg-Marquardt öğrenme algoritması ile 6 farklı kişinin konuşmaya-bağımlı ve konuşmaya-bağımsız konuşmalarının tanınmasında bir yaklaşım sunmuşlardır (Grieder ve Kinsner, 1994; Wong ve Sridharan, 2001; Campbell ve ark., 2006). Bu çalışmalarında %88 ile %94 oranında bir başarı sağlamışlardır.

Ergezer ve ark. (2003) Yapay Sinir Ağları ve Tanıma Sistemleri isimli çalışmalarında tanıma sistemleri hakkında kısa bir bilgi vererek yapay zeka üzerine yapılabilecek bu alandaki çalışmaların günümüz teknolojinin gelişimine büyük katkı sağlayacağını dile getirmiştir (Hermansky, 1990; Benesty ve ark., 2008).

Ses ile Kimlik Doğrulama isimli Yüksek Lisans Semineri çalışmasında ses ile kimlik doğrulama konusu ele alınmış ve ses sinyalinin üretilişi ve modellenmesi anlatılmıştır (Grieder ve Kinsner, 1994; Benesty ve ark., 2008). Ayrıca sistem performansının değerlendirilmesi, yanlış onaylama ve yanlış reddetme hataları ele alınmış ve sistem performansını etkileyen faktörler incelenerek mevcut ses tanıma sistemleri, insanlardaki ses tanıma sistemi ile karşılaştırılmıştır. (Zhi ve ark., 2001).

Speech Emotion Recognition Using Hidden Markov Models” isimli çalışmada hisli konuşma tanıma çalışması yapılmıştır (Borden ve Wilkins, 1984; Zhi ve ark., 2001). Bu çalışmada tanıma verimliliğini arttırmak için Gizli Markov Modeli kullanılmıştır. HMM-Based Channel Error Mitigation and its Application to Distributed Speech Recognition” isimli çalışmada, dağıtılmış konuşma tanımı için kanal hatasının azaltılmasını Temel Gizli Markov Modeli kullanılarak yapılması önerilmiş ve HMM'ye bağlı teknikler etkin bir şekilde kanal hatalarını azaltabileceği belirtilmiştir (Denes ve Pinson, 1963; Olson, 1972; Borden ve Wilkins, 1984).

High Quality Speech Transformation Based on Linear Prediction Coding and Pitch Synchronization isimli çalışmada, konuşma transformasyonu metodu olarak pitch eşlemesi ile birlikte Linear Prediction Code (LPC) tekniği tavsiye edilmiştir ve konuşma, farklı ses renkleri ve tonları ya da yüksek kalitede konuşma hızı için değiştirilebileceği ifade edilmiştir (Grieder ve Kinsner, 1994; Campbell ve ark., 2006; Benesty ve ark., 2008).

Yapay Sinir Ağları Modeli ile Ses Tanıma, farklı YSA algoritmaların ses tanımadaki performanslarının mukayesesi yapılarak ses tanımaya uygun olan algoritmanın belirlenmesi amaçlanmıştır. Ayrıca, bu çalışma ile YSA'lar ile ses tanıma problemine farklı bir yaklaşım getirilmiştir. Bu yaklaşım, YSA'nın da eğitilen ses komut verilerinden elde edilen sonuçlarla, tanıma ve kontrol işlemini gerçekleştirmek için, hata analizi yapılarak elde edilen hata değerlerini kullanmayı hedeflemiştir (Hermansky, 1990; Slaney, 1998).

Yukarıda, ses tanıma, konuşmacı tanıma, biyometrik sistemler, ses ile kimlik doğrulama, hisli konuşma tanıma, LPC Tekniğinin birkaç uygulaması ve YSA'ların hem ses tanıma hem de diğer mühendislik uygulamaları kısaca özetlenerek verilmiştir. Bugüne kadar yapılan çalışmalarda, ses tanıma probleminde Hopfield, Self Organizing Map (SOM), Radial Basis Function (RBF) gibi ağların ve Çok Katmanlı Perseptron (ÇKP) ağlarına ait algoritmaların kullanıldığı belirlenmiştir (Hermansky, 1990; Slaney, 1998; Bolia ve ark., 2000).

3. MATERYAL VE YÖNTEM

3.1. Ses Nedir?

Ses, titreşimli bir nesne tarafından oluşturulan bir basınç dalgasıdır. Bu titreşimler ortamdaki parçacıkları (tipik hava) titreşim hareketinde bulunmasına sebep olur, böylece enerjiyi ortamdan aktarır. Parçacıklar dalga hareketine paralel yönde hareket ettiğinden, ses dalgasına uzunlamasına dalga denir. Ses fizyolojisi ile ilgili bazı önemli kavramlar şöyle açıklanabilir.

Ses dalgası, ses, sıvı, katı ve gaz ortamlarında insan kulağı tarafından 20 Hz ve 20 KHz arasında tespit edilebilen basınç değişiklikleri olarak tanımlanır. Bu frekans aralığındaki mekanik dalgalar işitmemizi sağladığından, bizim için özellikle önemli olan ses dalgalarıdır. İnsan kulağına bir ses dalgası geldiğinde, kulak, ses dalgasındaki basınç değişikliklerini beyinde duyulan sesler olarak yorumlanan sinirlerdeki dürtülere dönüştürür. (Fukushima, 1975). En basit ses dalgasının sadece bir frekansı ve sabit bir genliği vardır. Buna sinüs dalgası denir. Frekans iki tepe arasındaki mesafe dalga boyudur. Saniyede gözlenen dalga zirvelerinin sayısı frekans olarak adlandırılır. Frekans, sesin perdesini belirler. Saniyede döngüler (CPS) veya Hertz (Hz) olarak görüntülenir. Düşük frekanslar bas sesleridir ve yüksek frekanslar tiz seslerdir (Haykin, 1999a; Farooq ve Datta, 2001). Genlik, sesin bir başka karakteristik özelliğidir. Bu, havayı sıkıştırmak için kullanılan kuvvete bağlı olarak havadaki küçük veya büyük bir basınca karşılık gelir (Haykin, 1999b; Farooq ve Datta, 2001).

3.1.1. Ses hangi ortamlarda yayılır?

Ses katı, sıvı ve gaz ortamlarda yayılır. (Honig ve ark., 2005; Krishnan ve ark., 2008). Ancak ses uzayda ve boşlukta yayılmaz. Örneğin: uzayda oluşan büyük patlamalarda ses duyulmaz.

3.1.2. Ses neden boşlukta yayılmaz?

Dalgalar genellikle iki ana gruba ayrılır: mekanik ve elektromanyetik dalgalar: elektromanyetik dalgalar yaymak için bir ortama ihtiyaç duymaz ve uzayda yayılabilir (Denes ve Pinson, 1963; Olson, 1972). Öte yandan mekanik dalgalar, enerjilerini aktarmak için ortam parçacıklarına ihtiyaç duyarlar. Bu nedenle, (örneğin uzayda) yayılamazlar (Zhu ve ark. 2001).

Ses dalgaları da mekanik dalgalar olduğundan, yaymak için maddi bir ortama ihtiyaç duyarlar. Ses, uygun bir ortamda (hava, su vb.) hareket eden nesnelerin titreşim dalgasıdır. Bir yerden diğerine sıkıştırma ve genişleme şeklinde yayılır. Bu nedenle, ses bir basınç dalgasıdır (Slaney, 1998; Bolia ve ark., 2000).

3.1.3. Sesin yayılma hızı

Sesin yayılma hızı, havada 20°C'de 340m/sn dir. Ses hızı ortam sıcaklığından etkilenir. Ses hızı frekansa bağlı olarak değişmez, her frekansta ses aynı hızda gider. Sıcaklıktaki artış moleküllerin titreşim hızını artırır, böylece sesin yayılma hızını artırır (Grieder ve Kinsner, 1994; Benesty ve ark., 2008).

3.1.4. Konuşma tanıma

Konuşma tanıma, insan sesinin bilgisayar tarafından algılanması olarak adlandırılmıştır. Ses tanıma alanındaki konuşma tanıma disiplini, teknoloji geliştirme sürecinde önemli bir yer edinmeye çalışan bir sistemdir ve bir bilgisayar tarafından bir mikrofon aracılığıyla insan sesini tanıma sürecidir. (Wong ve Sridharan, 2001; Kraetzer ve ark., 2007).

Konuşma tanıma için ilk uygulamalar arasında otomatik telefon sistemleri ve tıbbi dikte yazılımı vardır. Özellikle uzman sözlüklere dayanan mesleklerde, dikte için, veri tabanlarını sorgulamak ve bilgisayar tabanlı sistemlere komutlar vermek için sıklıkla kullanılmıştır. Herhangi bir makine konuşmayı yorumlayabilmeden önce, bir mikrofon ile bir kişinin sesinin titreşimlerini dalgalı bir elektrik sinyaline dönüştürmelidir. Bu sinyal, sistemin donanımı tarafından (örneğin, bilgisayarın ses

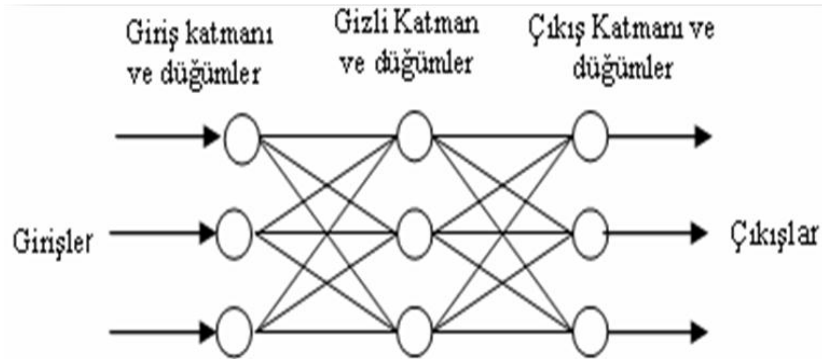
kartı) dijital bir sinyale dönüştürülür. Bir konuşma tanıma programının, konuşmanın temel yapı taşları olan ayrı fonemleri tanımak için analiz ettiği dijital sinyaldir. Fonemler daha sonra kelimelerle birleştirilir. Bununla birlikte, birçok kelime aynı şekilde ses çıkarır ve uygun sözcüğü seçmek için programın bağlamına dayanması gerekir. Birçok program, olasılıkları verilen herhangi bir iki sözcüğün ardından üçüncü bir kelimeyle takip edileceği sık sık üç kelimeli kümelerin bir veri tabanına dayanan bir yöntem olan trigram analizi yoluyla bağlam oluşturur. Örneğin, bir konuşmacı “kim olduğumu” söylerse, bir sonraki kelime, benzer sesten ziyade daha az “göz” yerine zamir “Ben” olarak tanınacaktır, ancak hataları düzeltmek için bazen insan müdahalesine ihtiyaç duyulabilmektedir.

3.1.4.1.Sesin kaydedilmesi ve ifadenin saptanması

İlk aşama, ‘Sesin kaydedilmesi ve ifadenin saptanması’ şeklindedir. Bu aşamada, ses tanıma sisteminin yapısına göre mikrofon veya telefon ile kayıt işlemi gerçekleştirilir. Sonrasında ‘Konuşma başlangıç ve bitişinin belirlenmesi’ gerçekleştirilir. Bu tez'de üst mikrofon ve kulaklık gerekmektedir ve *MATLAB* R2014b kullanılmıştır (Wong ve Sridharan, 2001; Campbell ve ark., 2006; Kraetzer ve ark., 2007).

3.2. Yapay Sinir Ağları (YSA)

Yapay zekâ (YZ), bir bilgisayarın ya da bilgisayar denetimli bir makinenin, genellikle insana özgü nitelikler olduğu varsayılan akıl yürütme, tanıma, tanımlama, sınıflandırma, konuşma, görme ve, genelleme ve geçmiş deneyimlerden öğrenme gibi yüksek zihinsel süreçlere ilişkin görevleri yerine getirme yeteneği olarak tanımlanmaktadır (Grieder ve Kinsner, 1994; Benesty ve ark., 2008).



Şekil 3.1. Yapay Sinir Ağları

3.2.1. Yapay sinir ağlarının sınıflandırılması

Danışmanlı öğrenme sırasında ağa verilen giriş değerleri için çıktı değerleri de verilir. Ağ verilen girdiler için istenen çıkışları oluşturabilmek için kendi ağırlıklarını günceller. Ağın çıktıları ile beklenen çıktılar arasındaki hata hesaplanarak ağın yeni ağırlıkları bu hata payına göre düzenlenir. Genel olarak YSA' lar, birbirleri ile ilişkili sinir hücrelerinden oluşmaktadırlar. Ağın yapısını her bir sinir hücresi arasındaki bağlantıların yapısı belirler. Öğrenme algoritması, hedeflenen değeri elde etmek için bağlantıların nasıl değiştirildiğini belirler. Kullanılan öğrenme algoritmasına göre, hatayı sifıra indirgemek için ağın ağırlıkları değiştirilir. YSA' lar yapılarına göre sınıflandırılırlar (Farley ve Clark, 1954; Elmas, 2003).

3.2.2. Yapay sinir ağının türleri

Günümüzde, belirli amaçlarla ve değişik alanlarda kullanılmaya uygun birçok yapay sinir ağı modeli (Perceptron, Adaline, MLP, LVQ, Hopfield, Recurrent, SOM, ART ve PCA gibi) geliştirilmiştir. Yapay sinir ağları türleri 2'ye ayrılır:

Tek Katmanlı Sinir Ağları

- Perceptron
- Çok Katmanlı Sinir Ağları
- Hopfield Ağı
- Kohonen Özellik Haritası

3.2.3. Yapay sinir ağının avantajları

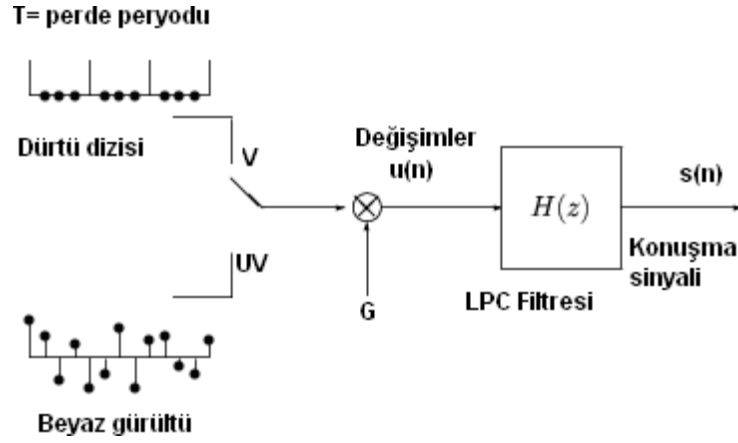
- YSA eğitim, öğrenme ve ilk deneyim için verilen verilere dayanarak nasıl görev yapılacağını öğrenme yeteneğine sahiptir.
- YSA kendi örgütlerini oluşturabilir ve herhangi bir denetim gerektirmez, kendi denetlenmeyen rekabetçi öğrenmelerini öğrenebilirler.
- YSA'nın hesaplamaları paralel olarak yapılabilir.
- YSA, verilerin kullanılması ve genelleştirilmesi için güçlü bir teknik olan örüntü tanımada kullanılabilir.
- Sistemin gelişimi programlama yerine öğrenme yoluyla gerçekleşir.
- YSA değişen ortamlarda esnekler.
- YSA'nın geleneksel modeli başarısız olduğunda bilgilendirici bir model oluşturabilir. Çok karmaşık etkileşimler yapabilirler.
- YSA, istatistiksel yöntemlerden kullanımı çok kolay ve anlaşılır olan doğrusal olmayan bir modeldir.

3.3. Doğrusal Öngörü Kodlaması (Linear Prediction Coding (LPC))

Doğrusal öngörüm kodlaması, ses sinyali ve konuşma sinyali işlemede en sık kullanılan yöntemlerden biri olmuştur. LPC genel olarak sinyallerin sıkıştırılmış halde bir gösterimi olarak da adlandırılır. LPC hem en güçlü konuşma analiz tekniklerinden birisi hem de iyi kalitede sıkıştırma sunabilen bir yöntemdir. Düşük bit sayısına sahip olduğu için yüksek kalitede sıkıştırma yapma imkânı vardır. LPC analizi sonucunda katsayılar ortaya çıkar iletim hattından gerekli yere iletilir daha sonra iletilen kısımdaki ters sistem ile yeniden orijinal sinyale çok yakın bir sinyal elde edilebilir.

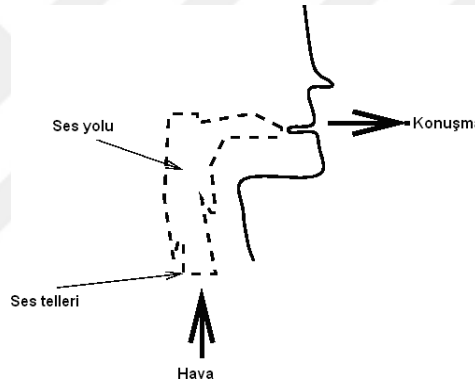
LPC çıktıları aynı zamanda çizgisel izge frekansları yönteminin de girdilerini oluştururlar bu da konuşma sinyalinin formatlarını tahmin etmeye yarar. Ayrıca LPC işlemi zaman bölgesinde yapılan bir analiz olarak adlandırılır.

Şekil 3.2'da ise LPC'nin konuşma sinyaline özgü olarak ses yolundaki modellenmesi gösterilmektedir.



Şekil 3.2. Konuşma Üretimi için Matematiksel Model

Bu model konuşma tanıma da kullanılabilir. Şekil 3.3'te LPC'nin genel fiziksel modeli gösterilmektedir.



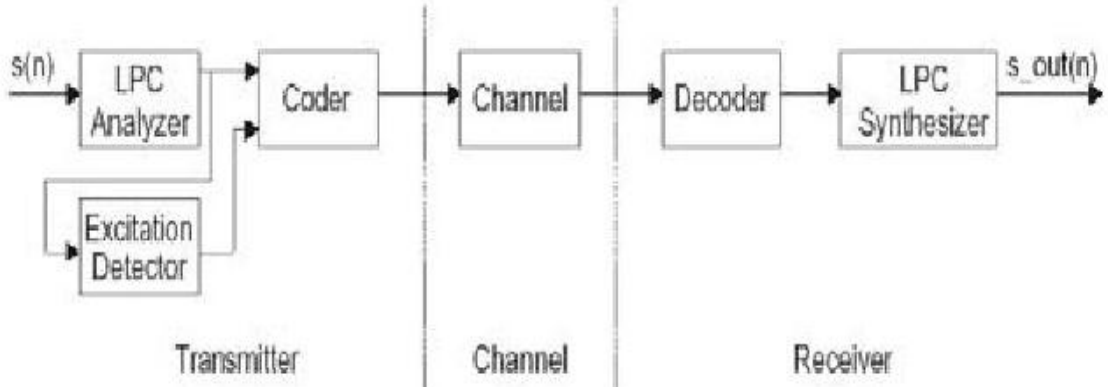
Şekil 3.3. Fiziksel Model

Bu bölümde LPC konuşma kodlama tekniğinin bir açıklaması verilecektir. Bazı sürtünme ve patlayıcı (veya kısaltılmamış) sesler için ses telleri titremez, sürekli açık kalır. Bazı seslerde ses telleri titrer. Ses tellerinin titreştiği hızda ses perdesini belirler. Ses yolunun şekli sesi belirler. Kişi konuşursa, ses yolu farklı ses üreterek şeklini değiştirir.

3.3.1. LPC katsayılarının ölçümü

Genellikle, doğrudan öngörüm katsayılarının nicel değerlendirilmesi göz önünde bulundurulmaz. Katsayıların stabilitesini sağlamak için (kutup ve sıfırların z -düzlemindeki birim daire içinde kalması gerekir) nispeten yüksek bir doğruluk

(katsayılar başına 8-10 bit) gereklidir. Bu, öngörüm katsayılarındaki küçük değişikliklerin kutup pozisyonlarında nispeten büyük değişikliklere yol açma etkisinden gelir. Bunlar iyi bilinen Levinson-Durbin özyinelemesinin hesaplanması sırasında kullanılan ara değerlerdir. Ara değerlerin nicelendirilmesi, öngörüm katsayılarının doğrudan ölçülmesinden daha az sorunludur.



Şekil 3.4. Sesli bir LPC ses kodlayıcısının blok şeması

LPC analizörünün tahmini transfer fonksiyonu ile filtrelenir. Bu filtrelenmiş artık sinyal olarak adlandırılır.

Her şeyden önce, iyi bir yeniden uyarma inşası için artık sinyal sadece düşük frekanslara ihtiyaç duyar. Yüksek bir sıkıştırma oranı elde etmek için artık sinyalin ayrık kosinüs dönüşümünü (DCT) kullanılmaktadır. DCT'nin, sinyalin enerjisinin çoğunu ilk birkaç katsayıda yoğunlaştırdığı bilinmektedir. Bu nedenle, sinyali sıkıştırmanın bir yolu, sadece enerjinin çoğunu içeren katsayıları transfer etmektir. Testlerimiz ve simülasyonlarımız bu katsayıların sadece 4 bit kullanılarak ölçülebileceğini bile göstermiştir. Alıcı sadece ters bir DCT yapar ve sesi uyarlamak için elde edilen sinyali kullanır.

3.3.2. LPC-motivasyon

LPC genelde yörüngesel zarf bilgilerini iletmeyi amaçlar. Filtre katsayılarının iletimi direkt olarak zordur bu nedenle daha ileri seviyede yöntemler geliştirilmiştir. LPC katsayıları kovaryans ya da otokorelasyon yöntemleri ile hesaplanabilir.

Förmül 3.1'de pth sipariş filtresi için konuşma farkı denklemi:

$$s(n) = \sum_{k=1}^p a_k s(n-k) + G u(n) \quad (3.1)$$

Ortalama kare tahmin hatasını en aza indirmek ister:

$$e(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (3.2)$$

Tek bir giriş dürtü veya sabit beyaz gürültü için, elde edilen katsayılar ak 's ile aynıdır.

3.3.3. LPC - Birinci otomatik korelasyon

Kod çözücü tarafından orijinal giriş sinyalini yeniden oluşturmak için kullanılan filtre, esas alınarak oluşturulur. Bu katsayılar kodlama sırasında orijinal sinyalden çıkarılır ve kod çözme işleminde kullanılmak üzere alıcıya iletilir.

S (n) 'nin $0 \leq n \leq N-1$ aralığı dışında sıfır olduğunu varsayarsak, Daha sonra aşağıdaki lineer denklemleri çözmemiz gerekir.

$$\sum_{k=1}^p a_k R(|i-k|) = R(i) \quad 1 \leq i \leq p \quad (3.3)$$

Nerede:

$$R(k) = \sum_{m=0}^{N-1-k} s(m)s(m+k) \quad (3.4)$$

LPC – İkinci otomatik korelasyon

Matris formunda lineer denklem seti şu şekilde ifade edilebilir:

$$\begin{pmatrix} r(0) & r(1) & \dots & r(M-2) & r(M-1) \\ r(1) & r(0) & \dots & r(M-3) & r(M-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r(M-2) & r(M-3) & \dots & r(0) & r(1) \\ r(M-1) & r(M-2) & \dots & r(1) & r(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_1 \\ \vdots \\ a_{M-1} \\ a_M \end{pmatrix} = \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(M-1) \\ r(M) \end{pmatrix} \quad (3.5)$$

3.3.4. LPC Uygulamaları

Genel olarak, konuşma sıkıştırma için en yaygın kullanım standart telefon sistemlerindedir. Aslında, konuşma sıkıştırmasında kullanılan teknolojinin çoğu telefon şirketleri tarafından geliştirilmiştir. Şekil 3.5, farklı telefon sistemleri tarafından kullanılan bit oranlarını göstermektedir. Doğrusal tahmin kodlaması, düşük bit hızı nedeniyle güvenli telefon alanında yalnızca 20 uygulamaya sahiptir. Güvenli telefon sistemleri, konuşma ilk önce dijitalleştirildikten sonra şifrelenip iletildiğinden düşük bit hızı gerektirir. Bu sistemler, anlaşılabilir bir konuşma kalitesi seviyesini korurken, bit hızını olabildiğince düşürmek gibi temel amaçlara sahiptir.

Kuzey Amerika Telefon Sistemleri	64 kb / s (sıkıştırılmamış)
Uluslararası Telefon Ağı	32 kb / s (5,3-64 kb / s arasında değişebilir)
Dijital Hücresel standartlar	6.7-13 kb/s
Bölgesel Hücresel standartlar	3.45-13 kb/s
Güvenli Telefon	0.8-16 kb/s

Şekil 3.5: Farklı telefon standartları için Bit Oranları

Anlaşılabilir konuşma kalitesi, Dijital hücresel standart ve diğer standartlar gibi uluslararası telefon şebekesi yüksek kalite standartlarına sahiptir ve bu nedenle daha yüksek bit hızı gerektirir. Bu standartlarda konuşmanın anlaşılması yeterince iyi değildir, dinleyicinin konuşmayı asıl kaynağa ait olarak tanıyabilmesi gerekir.

Doğrusal tahmin kodlamanın kullanıldığı ikinci bir alan Metin-Konuşma sentezidir. Bu tür bir sentezde konuşma metinden üretilmelidir. LPC sentezi, ses yolunun bir modeline dayanan konuşmanın oluşturulmasını içerdiğinden, metinden konuşma üretmek için mükemmel bir yöntem sağlar.

LPC ve diğer konuşma sıkıştırma şemalarının diğer uygulamaları sesli posta sistemleri, telefon yanıtıma makineleri ve multimedya uygulamalarıdır. Telefon uygulamalarından farklı olarak çoğu multimedya uygulaması tek yönlü iletişimi sağlar ve verilerin depolanmasını içerir. Konuşma içeren bir multimedya uygulaması örneği, bir metin belgesiyle ilgili sesli açıklamaların belge ile kaydedilmesine izin veren bir uygulamadır. Multimedya uygulamalarında kullanılan konuşma sıkıştırma yöntemi, istenen konuşma kalitesine ve uygulama için saklama alanının sınırlamalarına bağlıdır.

Doğrusal Öngörülü Kodlama, çoklu bit uygulamaları için uygun bir konuşma sıkıştırma yöntemi sağlar, çünkü düşük bit hızı nedeniyle en küçük depolama alanını sağlar.

3.4. Parçacık Sürü Optimizasyonu (PSO)

Kuş sürülerinin davranışlarından esinlenen J.Kennedy R.C ve Eberhart 1995 yılında, popülasyon tabanlı olan bu optimizasyon tekniğini geliştirmişlerdir (Kennedy ve Eberhart, 1995). Tasarlanmasındaki amaç doğrusal olmayan sorunları çözmektir. Bir diğer amacı ise, çok parametrelili ve çok değişkenli olan optimizasyon problemlerini çözmektir (Çevik ve Koçer, 2013).

Örneğin, $5 \times 2 + 2y^3 - (z / w)^2 + 4$ fonksiyonunun çözüm alanı, bilinmeyenler x, y, z ve w nedeniyle 4 boyutludur. Bu sorunun çözüm alanında tanımlanan bir parçacığın konumu, $P = [x, y, z, w]$ olarak 4 koordinat ile gösterilir.

Örneğin, yukarıdaki işlevi sıfıra eşitleyerek bir çözüm bulunabilir. $P = [-1, 0, 3, 1]$ parçacık için fonksiyon; $5 \times 2 + 2y^3 - (z / w)^2 + 4 = 0$ veya $P = [3, 3, 8, 1]$ şeklindeki bir parçacık $x = 3, y = 3, z = 8$ ve $w = 1$ koordinatlar için bir uygunluk fonksiyonunu gösterir. PSO için, insanların görsel olarak tasvir edemediği 4 veya daha fazla boyuttaki karmaşık problemleri incelemede zorluk yoktur.

Bir grup rastgele çözüm(parçacık sürüsü) ile PSO, başlatılır ve güncellemeler ile optimum çözümü bulmaya çalışır. Her yinelemede, parçacık konumları en iyi iki değere göre güncellenir. İlk olarak; parçacık tarafından elde edilen en iyi çözümü sağlayan koordinatlarıdır. "Pbest" adlı bu değer bellekte saklanmalıdır. Diğer en iyi değer, popülasyondaki tüm parçacıkların elde ettiği en iyi çözümü getiren koordinatlarıdır. Bu değer küresel değer olarak en iyisidir ve "gbest" ile gösterilir. Örneğin, D adet

parametreden oluşan n parçacıkları olduğunu varsayarsak. Bu durumda, popülasyon parçacık matrisi denklemi (3.6) gösterilen denklemde olduğu gibidir.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2D} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & \cdots & x_{nD} \end{bmatrix} \quad (3.6)$$

i ' inci parçacık formül 3.7 deki gibi ifade edilmektedir.

$$X_i = [X_{i1}, X_{i2}, \dots, X_{iD}] \quad (3.7)$$

i ' inci parçacığın önceki en iyi uygunluk değerini veren pozisyonu (pbest _{i});

$$pbest_i = [P_{i1}, P_{i2}, \dots, P_{iD}] \quad (3.8)$$

Formül 3.9'de ifade edilmektedir. gbest ise her iterasyonda tüm parçacıklar için tek olmalıdır ve

$$gbest_i = [P_1, P_2, \dots, P_D] \quad (3.9)$$

bu formülde gösterilmektedir. i ' ninci parçacığın hızı ise (her boyutundaki konumunun değişim miktarı)

$$V_i = [V_{i1}, V_{i2}, \dots, V_{iD}] \quad (3.10)$$

Förmül 3.11'de ifade edilmektedir. Parçacık hızları ve konumları iki en iyi değer bulunduktan sonra aşağıda gösterilen denkleme göre güncellenebilir.

$$V_i^{k+1} = v_i^k + c_1 \cdot rand_1^k \cdot (pbest_1^k - x_i^k) + c_2 \cdot rand_2^k (gbest^k - x_i^k) \quad (3.11)$$

$$p_i^{k+1} = p_i^k + v_i^{k+1} \quad (3.12)$$

(3.11) ' te gösterilen c_1 ve c_2 öğrenme faktörleridir. c_1 ve c_2 her parçacığı pbest ve gbest pozisyonlarına doğru çeker ayrıca hızlanma terimlerini ifade eden sabitler olarak adlandırılmaktadır.

3.4.1. Pbest- Gbest

İçgüdüsel olarak her parçacık iki şekilde hareket etme eğilimi göstermektedir. Bunlardan birincisi sürünün tamamı için en iyi pozisyon olan gbest'e yaklaşma eğilimi, diğeri ise, ve kendi en iyi pozisyonu olan pbest'i koruma eğilimidir. Sürüdeki parçacıkların pbest (personel best) değerleri aşağıdaki gibi belirlenir; ilk pbest, rastgele atanan başlangıç koşullarının değerlerine eşittir. Daha sonraki aşamalarda, önceki iterasyonda belirlenmiş pbest değeri ile karşılaştırılarak yeni bir pbest değeri atanır. Eğer t . iterasyondaki değeri $t+1$. iterasyondaki değerden fazlaysa iyi ise yeni pbest iterasyondaki değerdir.

$$\begin{cases} P_i(t+1) = x_i(t+1), & \text{eğer } f(x_i(t+1)) \leq f(P_i(t)), \\ P_i(t), & \text{değilse} \end{cases} \quad (3.13)$$

Sürünün gbest (global best) değeri 3.14' şekilde belirlenmektedir. Her bir iterasyon için pbest değerlerinden $F(x)$ fonksiyonunu minimum yapan değer, gbest değeridir. Her iterasyonda bu işlem tekrarlanır ve gbest güncellenir. Minimizasyon problemleri için :

$$p_g(t) = \arg_i \min f(P_i(t)) \quad (3.14)$$

Sürü içerisindeki parçacıkların tüm boyutlarını bulmak için pbestler ve gbestler ayrı ayrı hesaplanmaktadır.

3.4.2. Avantajları

- PSO'nun avantajlarından birisi reel sayılarla çalışıyor olmasıdır.
- Tasarım değişkenlerinin ölçeklendirilmesine duyarsızdır

- Uygulaması kolaydır.
- Eşzamanlı işleme için kolayca paralelleştirilebilir.
- Algoritma parametresi çok azdır
- Çok verimli gbest arama algoritması vardır
- Hız sınırlandırma

3.4.3. Dezavantajları

Hız sınırlandırmanın avantajları yanında dezavantajları da mevcuttur. Bunlardan ilki,

- hız sınırlaması problemi hız sadece adım büyüklüğünü değiştirmez, ayrıca kuşların hareket ettiği yönün de değişmesine neden olur
- Eylemsizlik ağırlığı(zayıf yerel arama yeteneği) mevcuttur.

3.4.4. Yöntemin uygulanışı

Başlangıç değerleri her parçacık için kısıtlamalara uygun olarak oluşturulur. Üretilen bu başlangıç değerleri ilk pbestleri oluşturacaktır. Öncelikle parçacığın (pbest) en iyi konumu belirlenmeli daha sonra da daha önce belirtilmiş olan kurallara ve işlemlere uygun olarak bir gbest belirlenmelidir. Komşuluk yönteminin kullanılıp kullanılmamasına bağlı olarak gbest hesabı değişiklik göstermektedir.

Konum vektörleri, atanan başlangıç değerlerine ve hesaplanan gbest ve pbest değerlerine göre belirlenmektedir. Parçacıkların hız vektörleri belirlenir. Normalde önceki yinelemeye (iterasyona) göre şekillendirilecek olan bu vektör, ilk yinelemede (iterasyonda) sıfır olacaktır. Başlangıçta parçacıkların duracağı varsayılmıştır. “c” ve “R” katsayılarının fonksiyon içerisinde kullanılması için belirlenmesi gerekmektedir. (bu katsayıların nasıl belirleneceği yukarıda anlatılmıştır). Her iterasyonda c1 ve c2 katsayıları sabitken, r1, r2 katsayıları rastgele belirlenmektedir.

Bu işlemler, belirli sayıda yineleme elde edilene kadar veya belirtilen bir kural yerine getirilene kadar devam eder. İşlemin tamamlanması durdurma kriterine bağlı olmalıdır. Durdurma kriterleri, yineleme sayısı veya en iyi sonuca yakınsama olarak

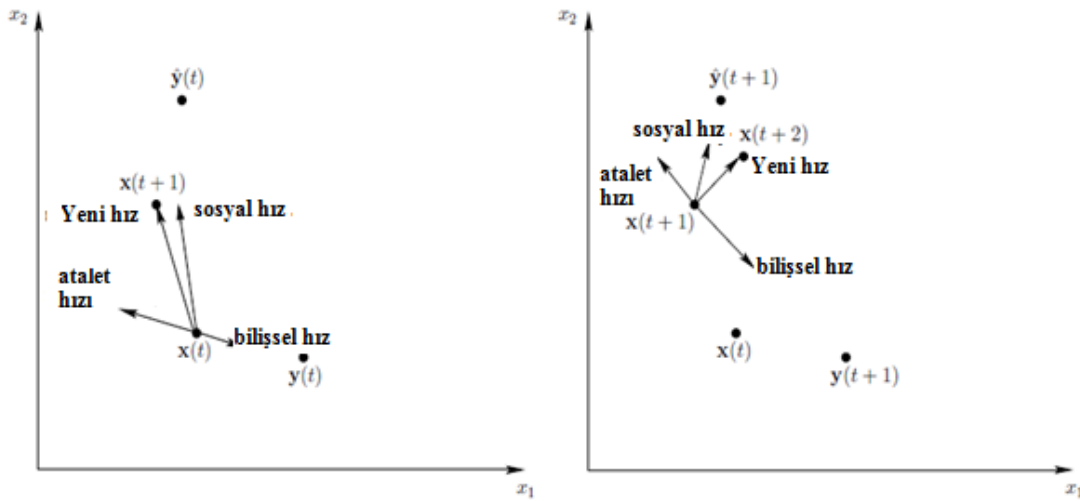
belirlenebilir. Yöntemi uygularken, sürüdeki parçacıkların sürüden ayrılma veya sapma gibi herhangi bir olasılığına karşı önlem alınmamaktadır. Çeşitli nedenlerden dolayı yoğun denklem sistemleri ve kısıtları olan sorunlarda bu durum sorunun çözümünü zorlaştırmaktadır. Sapmaları önlemek için çeşitli önlemler almak gerekmektedir.

3.4.5. Hız bileşenleri

(3.11) ve (3.12) denklemlerinde verilen hız hesaplaması üç terimden oluşur:

- **Önceki hız**, $v_i(t)$, önceki hareket ettiği yönüne göre işlev görür, yani yakın geçmişte hareket. Bu bellek terimi, parçacıkların sert bir şekilde yön değiştirmesini ve mevcut yöne doğru eğilmesini önleyen bir momentum olarak görülebilir. Bu bileşen aynı zamanda atalet bileşeni olarak da adlandırılır.

- **Bilişsel bileşen** olan $c_1 r_1 (y_i - x_i)$, i parçacığının performansını geçmiş performanslara göre ölçmektedir. Bir anlamda, bilişsel bileşen, parçacık için en iyi olan pozisyonun kişisel hafızasına benzer. Bu terimin etkisi, parçacıkları, bireylerin geçmişte kendilerini en çok tatmin eden durumlara veya yerlere dönme eğilimine benzer şekilde en iyi konumlarına çekilmeleridir. Kennedy ve Eberhart ayrıca bilişsel bileşene parçacığın “nostaljisi” olarak atıfta bulunmuştur.



(a) Zaman Basamağı t

(b) Zaman Basamağı $t + 1$

Şekil 3.7. Tek Bir İki Boyutlu Parçacık için Hız ve Konum Güncellemelerinin Geometrik Gösterimi.

Sosyal bileşen, sosyal bileşen parçacığının komşuluğunda bulunan en Kavramsal olarak iyi pozisyona çekilmesini sağlar, bireylerin elde etmeye çalıştığı bir grup normuna veya standardına benzemektedir.

Bilişsel ve sosyal bileşenlerin katkısı sırasıyla c_1r_1 veya c_2r_2 , stokastik bir miktar ile ölçülür.

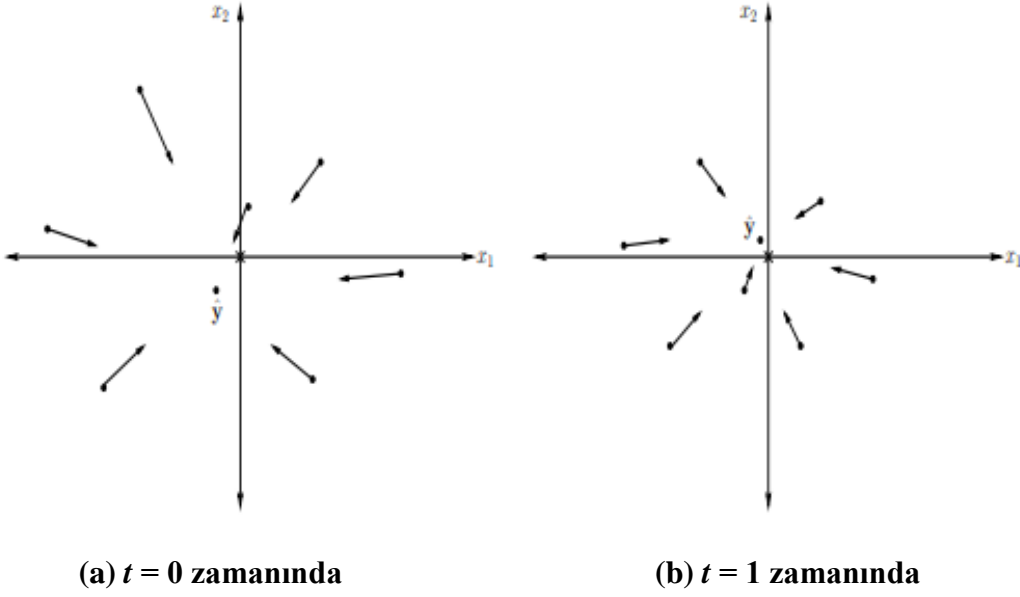
3.4.6. Parçacık sürü optimizasyonu ile geometri arama

Hız denkleminin etkisi, iki boyutlu bir vektör uzayında kolayca gösterilebilir. İki boyutlu bir uzayda tek bir parçacık göz önüne alındığında. Parçacığının örnek bir hareketi, şekil 3.7'de gösterilmiştir; burada parçacık alt simgesi, notasyonel kolaylık sağlamak için düşürülmüştür. Şekil 3.7 (a), t adımıdaki sürünün durumunu göstermektedir. Yeni konumun $x(t+1)$, global en iyi $y(t)$ 'ye doğru nasıl hareket ettiğine dikkat edildiğinde. $t+1$ adımı için, Şekil 3.7 (b) 'de gösterildiği gibi, kişisel en iyi pozisyonun değişmediği görülmektedir. Şekil, üç bileşenin hala parçacığı en iyi küresel parçacığa doğru nasıl hareket ettirmeye katkıda bulunduğunu göstermektedir.

Elbette, bir parçacığının, genel olarak momentum terimi nedeniyle, dünyanın en iyi pozisyonunu aşması mümkündür. Bu iki senaryo ile sonuçlanır:

1. Yeni pozisyon, mevcut küresel en iyiyi aştığından, mevcut küresel en iyisinden daha iyi bir pozisyon olabilir. Bu durumda, yeni parçacık pozisyonu yeni küresel en iyi pozisyon olacak ve tüm parçacıklar ona doğru çekilecektir.
2. Yeni pozisyon şu anda mevcut en iyi küresel parçacıktan daha kötü. İlerleyen zamanlardaki adımlarda, bilişsel ve sosyal bileşenler, parçacığının küresel en iyiye doğru yön değiştirmesine neden olacaktır.

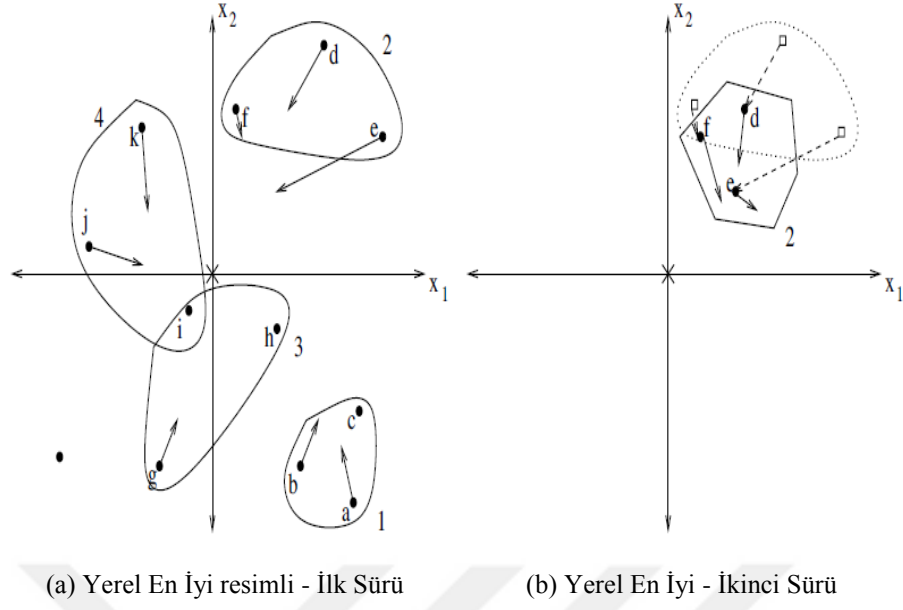
Bir parçacığının tüm konum güncellemelerinin kümülatif etkisi, her parçacığının hat üzerinde küresel en iyi pozisyonu ve parçanın en iyi kişisel pozisyonunu birleştiren bir noktaya yaklaşmasıdır.



Şekil 3.8 Çok Parçalı Gbest PSO Çizimi

Birkaç parçacık ve gbest Şekil 3.8'de gösterilmektedir. PSO'yu kullanarak x_1 ve x_2 değişkenleriyle iki boyutlu bir fonksiyonu en aza indirme görevine referansla pozisyon güncellemelerini görselleştirir. Optimum, "x" sembolü ile gösterilen başlangıç noktasıdır. Şekil 3.8 (a), belirtildiği gibi en iyi küresel pozisyonla birlikte sekiz partikülün ilk pozisyonunu göstermektedir. Bilişsel bileşenin katkısı $t = 0$ basamağında her parçacık için sıfır olduğundan, yalnızca sosyal bileşenin konum ayarlamaları üzerinde etkisi vardır. Küresel en iyi pozisyon değişmemektedir (tüm parçacıklar $\mathbf{v}_i(0) = \mathbf{0}$, için olduğu varsayılır). Şekil 3.8 (b), ilk yinelemeden sonra tüm parçacıkların yeni pozisyonlarını göstermektedir. Yeni bir küresel en iyi pozisyon bulunmuştur. Şekil 3.8 (b) şimdi, yeni küresel en iyi konuma doğru hareket eden parçacıklar ile tüm hız bileşenlerinin etkisini göstermektedir.

Son olarak, Şekil 3.9'te gösterildiği gibi PSO, parçacıkların acil komşularından nasıl etkilendiğini gösterir. Grafiği okunabilir kılmak için sadece hareketlerin bir kısmı gösterilmiştir ve sadece toplam hız yönü gösterilir. Bölge 1'de, hem a hem de b parçacıkları, bölge içindeki en iyi çözüm olan c parçasına doğru hareket eder. bölge 2 göz önüne alındığında, d parçacığı f 'ye doğru hareket eder; Bir sonraki yineleme için, e bölge 2 için en iyi çözüm olacaktır. Şimdi d ve f , Şekil 3.9 (b) 'de gösterildiği gibi e 'ye doğru hareket eder (çözüm alanının sadece bir kısmı gösterilmiştir). Bloklar önceki pozisyonları temsil eder e 'nin bölge 2 için en iyi çözüm olmaya devam ettiğine dikkat edin.



Şekil 3.9 lbest PSO'sunun çizimi

3.4.7. PSO 'nun tartışılan yönleri

Algoritmaların 3.11 ve 3.12 sayılı birkaç yönü hala tartışılmalıdır. Bu yönler arasında partikül başlatma, durma koşulları, yineleme ve fonksiyon değerlendirme terimlerinin tanımlanması yer almaktadır.

Algoritmalar 3.11 ve 3.12'ye atfen, optimizasyon işlemi yinelemelidir. Algoritmaların tekrarlanan yinelemeleri, durma koşulu sağlanıncaya kadar yürütülür. Bu tür bir yineleme, tekrarlama içindeki tüm adımların uygulanmasından sonuna kadar, yani kişisel en iyi pozisyonları ve en iyi küresel pozisyonu belirlemek ve her partikülün hızını ayarlamaktan oluşur. Her yinelemede, bir dizi fonksiyon değerlendirmesi (FEs) gerçekleştirilir. Bir *fonksiyon değerlendirmesi* (FE), optimizasyon problemini karakterize eden uygunluk fonksiyonunun bir hesaplamasını ifade eder. Temel PSO için, yineleme başına toplam n_s fonksiyon değerlendirmeleri yapılır; burada n_s , kümedeki toplam parçacık sayısıdır.

PSO algoritmasının ilk adımı sürü ve kontrol parametrelerini başlatmaktır. Temel PSO bağlamında, ivme sabitleri, c_1 ve c_2 , başlangıç hızları, parçacık pozisyonları ve kişisel en iyi pozisyonların belirtilmesi gerekir. Ek olarak PSO, bölge boyutunun belirtilmesini gerektirir. Genellikle, parçacıkların konumları, arama alanını eşit olarak

kapsayacak şekilde başlatılır. PSO'NUN verimliliğinin sürünün ilk çeşitliliğinden etkilendiğini, yani arama alanının ne kadarının kaplandığını ve parçacıkların arama alanına ne kadar iyi dağıtıldığını belirtmek önemlidir. Arama alanının bölgeleri başlangıç sürüsü tarafından karşılanmazsa, PSO, ortaya çıkarılan bir bölgede bulunuyorsa optimum bulmakta zorluk çekecektir. PSO, bir parçacığın momentumu, parçacığın kendisi için yeni bir kişisel en iyi ya da yeni küresel en iyi haline gelen bir pozisyona ulaşması koşuluyla, parçacığı ortaya çıkarılan alana taşırsa böyle bir optimum keşfedecektir.

Optimumun iki vektör tarafından tanımlanan etki alanı içinde bulunması gerektiğini varsayalım, x_{min} ve x_{max} , sırasıyla Her boyuttaki minimum ve maksimum aralıkları temsil eder. Daha sonra, parçacık konumları için verimli bir başlatma yöntemi: $x(0) = x_{min,j} + r_j (x_{max,j} - x_{min,j})$, $\forall j = 1, \dots, n_x$, $\forall i = 1, \dots, n_s$ nerede $r_j \sim U(0,1)$.

Başlangıç hızları sıfıra başlatılabilir, yani.

$$v_i(0) = \mathbf{0}$$

Hızları rastgele değerlerle de başlatmak mümkündür ama gerekli değildir ve bu özenle yapılmalıdır. Aslında, başlangıç konumlarındaki fiziksel nesnelere göz önüne alındığında, hızları sıfırdır – yani sabittirler. Parçacıklar sıfır olmayan hızlarla başlatılırsa, bu fiziksel benzetme ihlal edilir. Pozisyon vektörlerinin rastgele başlatılması zaten rastgele pozisyonlar ve hareketli yönler sağlar. Bununla birlikte, hızlar da rastgele başlatılırsa, bu hızlar çok büyük olmamalıdır. Büyük ilk büyük ilk hız, ivme, ve sonuç olarak büyük ilk pozisyon güncellemeler olacaktır. Bu büyük başlangıç konumu güncellemeleri, parçacıkların arama alanının sınırlarını terk etmesine neden olabilir ve parçacıkların tek bir çözüm üzerinde yerleşmeden önce sürünün daha fazla yineleme yapmasına neden olabilir.

Her partikül için en iyi kişisel pozisyon, $t = 0$ basamağında partikülün pozisyonunda başlatılır, yani $y_i(0) = x_i(0)$

3.5. Destek Vektör Makineleri (SVM)

Destek vektör makineleri (Support Vector Machine-SVM) sınıflandırma problemleri için kullanılır. Kullanımı basit ve son derece etkili farklı sınıflandırma yöntemlerindedir.

Varsayalım ki, $i = 1, 2, \dots, N$ olan, $x_i \in \mathbb{R}^n$ olan bir dizi S verilmişse, her bir nokta x_i , iki sınıftan birine aittir ve bu nedenle $y_i \in \{-1, 1\}$ etiketine sahiptir. Amaç, aynı sınıfın tüm noktalarını aynı tarafta bırakarak S 'yi bölen bir hiper düzlemin denklemini oluşturmaktır. SVM, verileri en iyi şekilde iki kategoriye ayıran bir N boyutlu oluşturarak sınıflandırmayı gerçekleştirir.

3.5.1. Verilerin sınıflandırılması

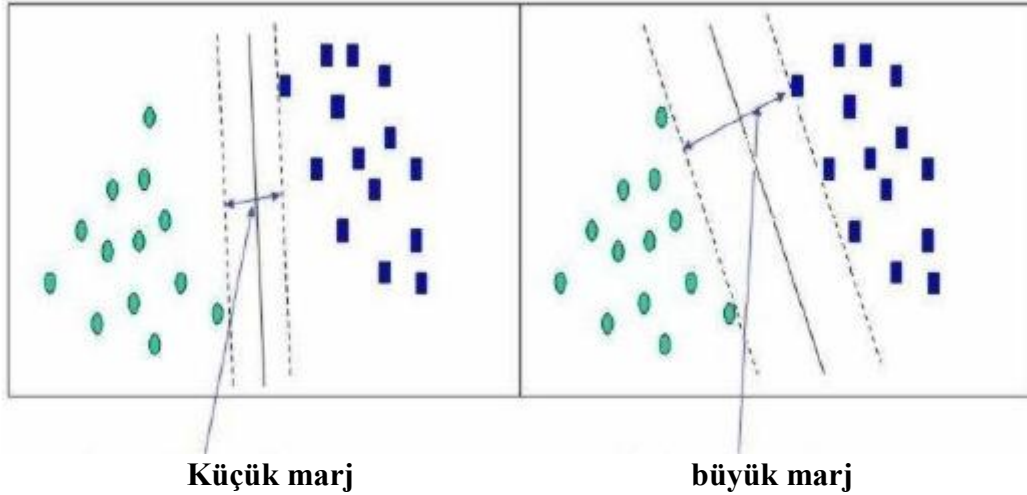
Nesnelerin iki farklı sınıfa ait olduğu görülür. İkinci resimdeki ayırma çizgisi (2 boyutlu hiper düzlem), nesnelere, her alt kümedeki tüm öğelerin benzer olacağı şekilde iki alt gruba ayıran bir karar düzlemdir.



Şekil 3.10. Veri Sınıflandırma

Not: Belirli bir nesne kümesi için birçok olası ayırma çizgisi vardır. Tüm ayırım çizgileri (karar sınırları = karar düzlemleri) eşit derecede olmaktadır.

Olası hiper düzlemler arasında, hiper düzlemin en yakın veri noktalarından (“kenar boşluğu”) mümkün olduğu kadar büyük olduğu nokta seçilir. Hiper düzlemi herhangi bir veriden mümkün olduğunca uzağa getirerek, r 'nin uygun şekilde büyük olmasına izin verir. İstenen hiper düzlem (marjı maksimuma çıkararak) aynı zamanda iki veri setinin dışbükey gövdesi üzerindeki en yakın noktalar arasındaki çizginin kesicisidir.



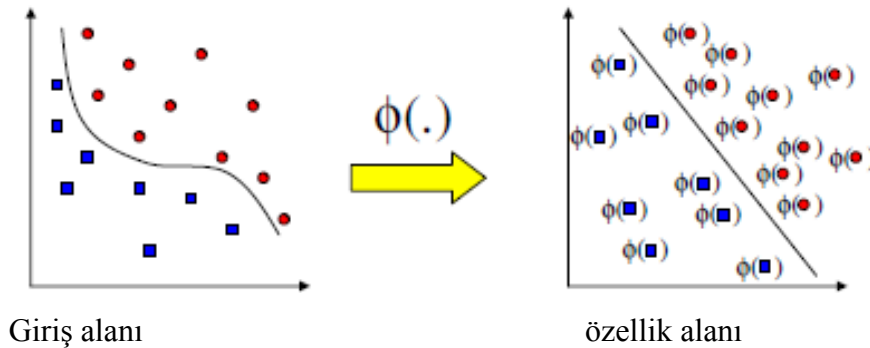
Şekil 3.11. Küçük ve büyük marjlar örneği

3.5.2. Verileri dönüştürmek

İki sınıf arasındaki ayırma sınırını tanımlayan matematiksel denklem basit olmalıdır. Bu yüzden giriş alanının verileri özellik alanına göre yeniden düzenlenir. Eşleme (yeniden düzenleme), özellik alanının artan boyutunu içerir. Veri noktaları, eğitim için veya sınıflandırma için kullanılmadan önce giriş alanından yeni bir özellik alanına eşleştirilir.

Verileri dönüştürdükten ve öğrendikten sonra daha basit özellik alanını incelenerek bir cevap aranır.

$$x = (x_1, \dots, x_N) \mapsto \phi(x) = (\phi_1(x), \dots, \phi_N(x)) \quad (3.15)$$



Şekil 3.12. Giriş alanını ve özellik alanını göstermektedir

3.5.3. Öğrenme

Öğrenme, iki uç nokta arasındaki hiper düzlemi ayıran maksimum marjı bulmak olarak kabul edilmektedir. Bir çiftin (w, b) , denklem 3.19 da ifade ettiği gibi içeren bir hiper düzlemi tanımladığını varsayarsak:

$$w \cdot x + b = 0 \quad (3.16)$$

$\{x_1, \dots, x_m\}$ veri kümemiz olsun ve $y_i \in \{1, -1\}$ x_i 'nin sınıf etiketi olsun. Karar sınırı tüm noktaları doğru bir şekilde sınıflandırmalıdır, yani aşağıdaki denklemlerin yerine getirilmesi gerekir:

$$w \cdot x_i + b \geq 1 \text{ eğer } y_i = 1 \quad (3.17)$$

$$\rightarrow y_i(w \cdot x_i + b) \geq 1$$

$$w \cdot x_i + b \leq -1 \text{ eğer } y_i = -1 \quad (3.18)$$

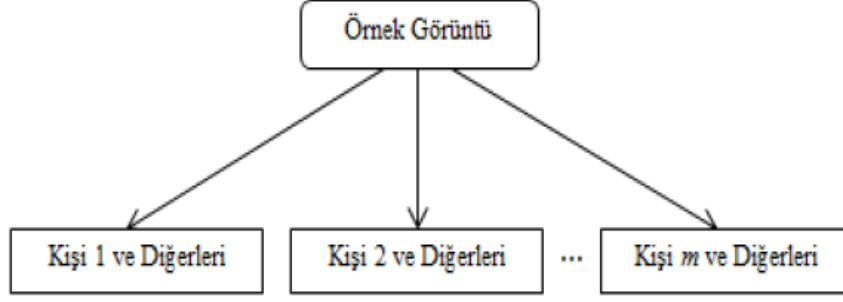
Verileri ayıran tüm hiper düzlemler arasında, aşağıdaki şekilde tespit edilebilecek sınıflar arasındaki maksimum ayrılma marjını veren benzersiz bir tane vardır;

$$\max \min \{ \|x - x_i\| : x \in \mathbb{R}^N, (w \cdot x) + b = 0, i = 1, \dots, m \} w, b \quad (3.19)$$

3.5.4. Çok sınıflı destek vektör makineleri

Destek vektör makineleri yöntemini ve diğer sınıflandırma yöntemlerini ayıran şey, iki sınıf arasında ayırım yapma yeteneğidir. SVM yöntemini ikiden fazla sınıfa sahip bir sistemde kullanmak için çeşitli algoritmalar vardır. Bu algoritmaların en çok kullanılanları bire karşı hepsi ve bire karşı bir algoritmalarıdır. Bu çalışmada kullanılan ise, SVM ile gerçekleştiren bire karşı hepsi algoritmasıdır.

Eğitim aşamasındaki sınıflardan biri tek bir yönteme karşı (+) olarak kabul edilirken, kalan tüm sınıfların bilgileri (-) olarak kabul edilir ve M sınıfı için Madet SVM oluşturulur. Tanınacak örnek, SVM tarafından (+) sınıflandırıldığında o sınıfa ait olmaktadır (Heisele ve ark., 2003). Bire karşı hepsi yönteminin örnek bir düzenlemesi, şekil 3.9'te gösterilmiştir.



Şekil 3.13. SVM için bire karşı hepsi örneği

3.5.5. Destek vektör makinelerin avantajları ve dezavantajları

Destek vektör makinelerinin çok fazla kullanılmasının nedeni: ilk olarak öğrenme, basit fikirler üzerine kurulmuş olma ve pratik uygulamalarda yüksek performans gösterebilmesidir. Kullanılacak olan örnek sayısı SVM'lerde önemli olmamaktadır. SVM eğitim anında daha önce görülmemiş verileri de hiçbir sıkıntı yaşamadan sınıflandırabilmektedir. Bu durumda SVM'nin genelleştirebilme yeteneğini gösterir. SVM'nin diğer tekniklere nazaran iyi bir alternatif olmasını sağlayan genelleştirebilme özelliği olmasıdır. Günümüzde SVM, veri madenciliğinde, örüntü tanınmasında, yüz bulma ve tanımada, dil yapısını inceleyen mantıksal programlamalarda, uçak alt basınç profillerinin modellenmesinde, biyoloji ve diğer biyo informatik uygulamalarda gen analizlerinde ve proteinlerin sınıflandırılmasında kullanılmaktadır(Kecman, 2001). Bazı durumlarda Lagrange Optimizasyon fonksiyonunun hesaplanması sorunlu ve karmaşık bir hal alabilmektedir.

Matematiksel olarak basit olarak ifade edilmesine rağmen, polinom derecesindeki artış algoritmanın karmaşık hale gelmesine neden olabilir. Bu işlem süresini önemli ölçüde artırmakla kalmaz, aynı zamanda bir noktada sınıflandırmanın doğruluğunu da azaltır. Bu algoritma için bir dezavantaj sağlar. Öte yandan, genelleme iyi çalışıyor olsa da, veri çiftlerini eğitmek için en iyi model olmayabilir. Algoritmanın bir diğer dezavantajı, nitelikleri istatistiksel olarak modelleyen ve doğrudan yüz görüntülerinde çalışırsa başarısız olan bir öğrenme modeli olmasıdır (Hsu ve ark., 2003).

3.5.6. Yumuşak marj sınıflandırıcı

Gerçek dünya probleminde, verileri alan içinde tam olarak bölen ayrı bir çizgi elde etme olasılığı düşüktür. Ve eğri bir karar sınırı olabilir. Verileri tam olarak ayırabilecek bir hiperplane olabilir, ancak verilerin içinde gürültü varsa bu arzu edilmeyebilir. Pürüzsüz sınırın, birkaç veri noktasını görmezden gelmesi, kavisli olmaktan veya aykırı değerlerin etrafında döngülere girmekten daha iyidir. Bu farklı bir şekilde ele alınır; burada gevşek değişkenler teriminin tanıtıldığı görülmektedir. Şimdi elimizde, $y_i (w'x_i + b) \geq 1 - S_k$ (Burges C.,1998),(J.P.Lewis,2004). Bu, bir noktanın kısıtlamayı ihlal etmeden hiper düzlemin yanlış tarafında küçük bir mesafe olmasını sağlar. Şimdi, herhangi bir çizginin verileri ayırmasına izin veren büyük durgunluk değişkenlerine sahip olabilir, bu nedenle bu gibi senaryolarda, büyük durgunluklara ceza veren Lagrangian değişkeni vardır.

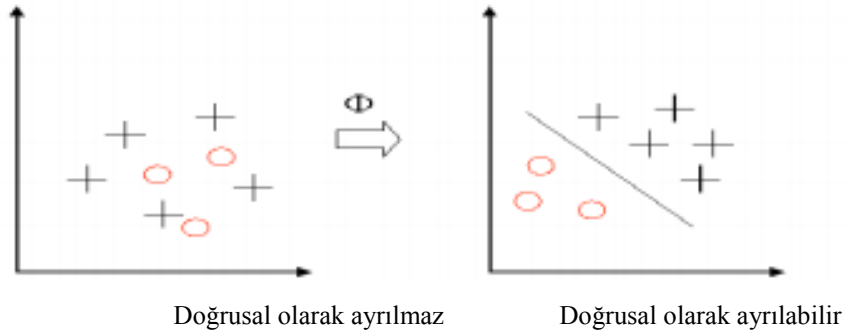
$$\min L = \frac{1}{2} w'w - \sum \lambda_i (y_i (w'x_i + b) + s_i - 1) + \alpha \sum s_i \quad (3.20)$$

Burada α azaltılması hiper düzlemin yanlış tarafında daha fazla veri sağlar ve daha düzgün karar sınırı vermek aykırı olarak kabul edilir(J.P.Lewis,2004).

3.5.7. Çekirdek numarası

Öncelikle, çekirdek nedir ve özellik alanı ne anlama geliyor?

Çekirdek: Veriler doğrusal ise, verileri bölmek için ayırıcı bir hiper düzlem kullanılabilir. Bununla birlikte, genellikle verilerin doğrusal olmadığı ve veri kümelerinin ayrıştırılmadığı durumlar söz konusudur. Bu çekirdeğe izin vermek için girdi verilerini yüksek boyutlu bir uzaya doğrusal olmayan bir şekilde eşlemek için kullanılır. Yeni haritalama daha sonra doğrusal olarak ayrılabilir. Bunun çok basit bir gösterimi Şekil 7'de aşağıda gösterilmiştir (Tom Mitchell,1997).

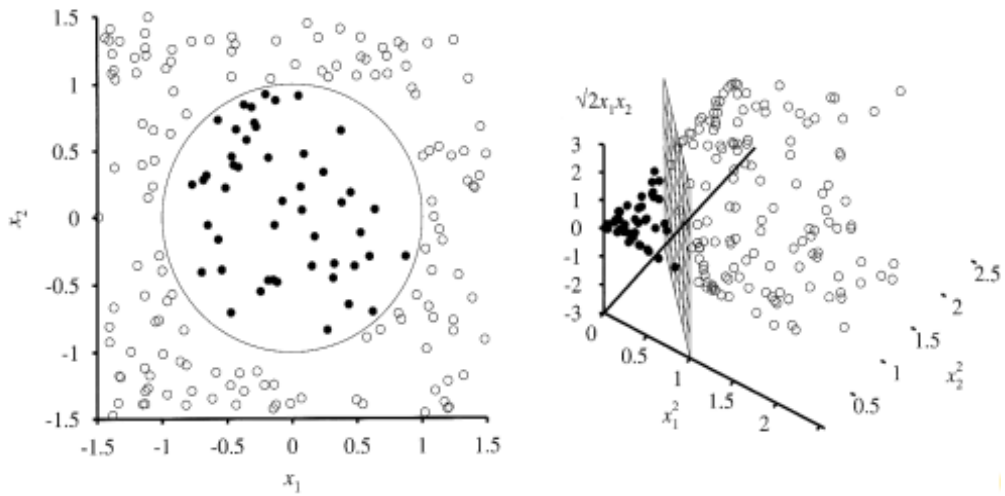


Şekil 3.14: Neden Çekirdek kullanılıyor

Bu haritalama, Çekirdek tarafından tanımlanır: $K(x, y) = \phi(x) \cdot \phi(y)$ (3.21)

Verilerin özellik alanına dönüştürülmesi, nokta ürün bazında benzerlik ölçüsü tanımlamayı mümkün kılar. Özellik alanı uygun şekilde seçilirse, desen tanıma kolay olabilir.

$$\langle x \cdot x \rangle \leftarrow K(x, x) = \langle \phi(x) \cdot \phi(x) \rangle \quad (3.22)$$



Şekil 3.15: Özellik Alanı Gösterimi

Not: İlgili kavramları anlamak için örnek komplo olduklarından açıklanmadığını unutmayın.

Çekirdek numarasına geri dönüldüğünde, w , b elde edildiğinde sorunun bir hiper düzlemle ayrıldığı basit bir doğrusal senaryo için çözüldüğü görülmektedir. Kenral numarası, SVM'lerin doğrusal olmayan sınırlar oluşturmalarına izin vermektedir. Çekirdek numarasıyla ilgili adımlar aşağıda verilmiştir (J.P.Lewis,2004),(Burges B. 1998).

[1] Algoritma, yalnızca veri kümelerinin iç ürünlerini kullanarak ifade edilir. Buna ikili problem de denir.

[2] Her veri vektörüne orijinal veri boyutunun bir çiftinin akıllı ürününü ekleyerek yeni boyutlara ilişkin yeni veriler oluşturmak için orijinal veriler doğrusal olmayan haritalardan geçirilir.

[3] Bu yeni, daha büyük vektörlerdeki bir iç üründen ziyade tablolarda saklanır ve daha sonra bir tablo araması yapılır, doğrusal olmayan haritalama yaptıktan sonra verilerin bir nokta ürününü temsil edebilir. Bu işlev çekirdek işlevidir. Çekirdek fonksiyonları hakkında daha fazla bilgi aşağıda verilmiştir.

3.5.7.1. Çekirdek Numarası: İkili Problem

İlk önce, optimizasyonla ilgili problem olan w 'yi ortadan kaldırmaya çalıştığımız ikili forma dönüştürürüz. Bir Lagrangian sadece λ_i 'nin bir fonksiyonudur. Bunun için matematiksel bir çözüm var, ancak bu öğreticinin matematiksel denklemleri en aza indirmek için talimatları vardır. Sorunu çözmek için LD'yi λ_i 'ye göre maksimize etmeliyiz. İkili form, optimizasyonu basitleştirir ve ana başarının bundan elde edilen nokta ürünü olduğu görülmektedir.

3.5.7.2. Çekirdek Numarası: İç Ürün Özeti

Burada, kullanılan veri vektörlerinin nokta ürününü temsil etmesi gerektiğini görülmektedir. Doğrusal olmayan haritalanmış verilerin nokta ürünü pahalı olabilir. Çekirdek numarası, doğrusal olmayan eşlemenin nokta ürününe karşılık gelen uygun bir işlevi seçilmektedir (Borges C.,1998), (Nello Cristianini ve John Shawe-Taylor,2000), (J.P.Lewis,2004). En sık seçilen çekirdek işlevlerinden bazıları, bu eğitimin sonraki kısımlarında aşağıda verilmiştir. Belirli bir çekirdek yalnızca deneme setinde deneme yanılma yöntemiyle seçilir, soruna veya uygulamaya göre doğru çekirdeğin seçilmesi SVM'nin performansını artırır.

3.5.8. Çekirdek İşlevleri

Çekirdek işlevi fikri, potansiyel olarak yüksek boyutlu özellik alanından ziyade giriş alanında işlemlerin gerçekleştirilmesini sağlamaktır. Bu nedenle, iç ürünün özellik alanında değerlendirilmesi gerekmez. İşlevin giriş alanının niteliklerinin özellik alanına eşlenmesini isteriz. Çekirdek işlevi SVM'de ve performansında kritik bir rol oynar. Çekirdek Hilbert Uzaylarının çoğaltılmasına dayanır.

$$K(x, x') = \langle \phi(x), \phi(x') \rangle, \quad (3.23)$$

K , Mercer'ın koşullarını karşılayan simetrik pozitif bir kesin fonksiyon ise,

$$K(x, x') = \sum_m^{\infty} a_m \phi_m(x) \phi_m(x'), \quad a_m \geq 0, \quad (3.24)$$

$$\iint K(x, x') g(x) g(x') dx dx' > 0, \quad g \in L^2 \quad (3.25)$$

Daha sonra çekirdek, özellik alanındaki meşru bir iç ürünü temsil eder. Eğitim seti bir giriş alanından doğrusal olarak ayrılamaz ama Eğitim seti, özellik alanından doğrusal olarak ayrılabilir. Buna da “Çekirdek numarası” denir (Nello Cristianini ve John Shawe-Taylor,2000), (J.P.Lewis,2004).

3.5.9. Regresyon için SVM

Svm'ler, alternatif bir kayıp fonksiyonunun getirilmesiyle regresyon problemlerine de uygulanabilir (Nello Cristianini ve John Shawe-Taylor,2000),(A. J. Smola.,1996). Kayıp fonksiyonu bir mesafe ölçüsü içerecek şekilde değiştirilmelidir. Regresyon doğrusal ve doğrusal olmayan olabilir. Doğrusal modeller esas olarak e-yoğun kayıp fonksiyonları, kuadratik ve Huber kayıp fonksiyonlarından oluşur. Sınıflandırma problemlerine benzer şekilde, verileri yeterince modellemek için doğrusal olmayan bir model gereklidir. Doğrusal olmayan SVC yaklaşımı ile aynı şekilde, doğrusal olmayan bir haritalama, verileri doğrusal regresyonun gerçekleştirildiği Yüksek boyutlu bir özellik alanına eşlemek için kullanılabilir. Çekirdek yaklaşımı yine boyutluluğun lanetini ele almak için kullanılır. Regresyon yönteminde, sorunun önceden bilgisine ve gürültünün dağılımına dayanan hususlar vardır. Bu tür bilgilerin yokluğunda Huber'ın sağlam kayıp fonksiyonu, iyi bir alternatif olduğu gösterilmiştir (Nello Cristianini ve John Shawe-Taylor,2000),(C. Cortes ve V. Vapnik.,1995).

3.5.10. Destek Vektör Makinelerinin Uygulamaları

SVM'nin örüntü sınıflandırma problemlerinde kullanıldığında başarılı olduğu görülmüştür. Destek Vektörü yaklaşımını belirli bir pratik probleme uygulamak, problem tanımına ve onunla ilgili tasarıma dayalı olarak bir dizi sorunun çözümlenmesini içerir. En büyük zorluklardan biri, verilen uygulama için uygun bir çekirdeğin seçilmesidir. Varsayılan seçenekler olan bir Gauss ya da polinom çekirdeği gibi standart seçenekler vardır, ancak bunlar etkisiz kalırsa ya da girdilerin ayrık yapılar olması halinde daha ayrıntılı çekirdeklere ihtiyaç duyulur. Çekirdek, bir özellik boşluğu tanımlayarak, makine tarafından verileri görüntülemek için kullanılan açıklama dilini sağlar. Çekirdek seçimi ve optimizasyon kriteri yapıldıktan sonra sistemin kilit bileşenleri aynı kalmaktadır (Borges C.,1998),(Nello Cristianini ve John Shawe-Taylor,2000).

Metin kategorizasyonun görevi, önceden tanımlanmış kategorilerin sabit bir sayısına ve doğal metin belgelerinin içeriğine göre sınıflandırılmasıdır. Bir belge birden fazla kategoriye atanabileceğinden, bu çok sınıflı bir sınıflandırma sorunu değildir, ancak her kategori için bir dizi ikili sınıflandırma sorunu olarak görülebilir. Bilgi alma amacıyla metnin standart temsillerinden biri, bir Mercer çekirdeği oluşturmak için ideal bir özellik eşlemesi sağlar. Gerçekten de, çekirdekler bir şekilde örnekler arasında bir benzerlik ölçüsü içerir ve belirli uygulama alanında çalışan uzmanların, özellikle bilgi alma ve üretken modeller gibi alanlarda geçerli benzerlik önlemlerini zaten belirlediklerini varsaymak mantıklıdır(Osuna E., Freund R., and Girosi F.,1997),(Veropoulos K., Cristianini N., and Campbell C.).

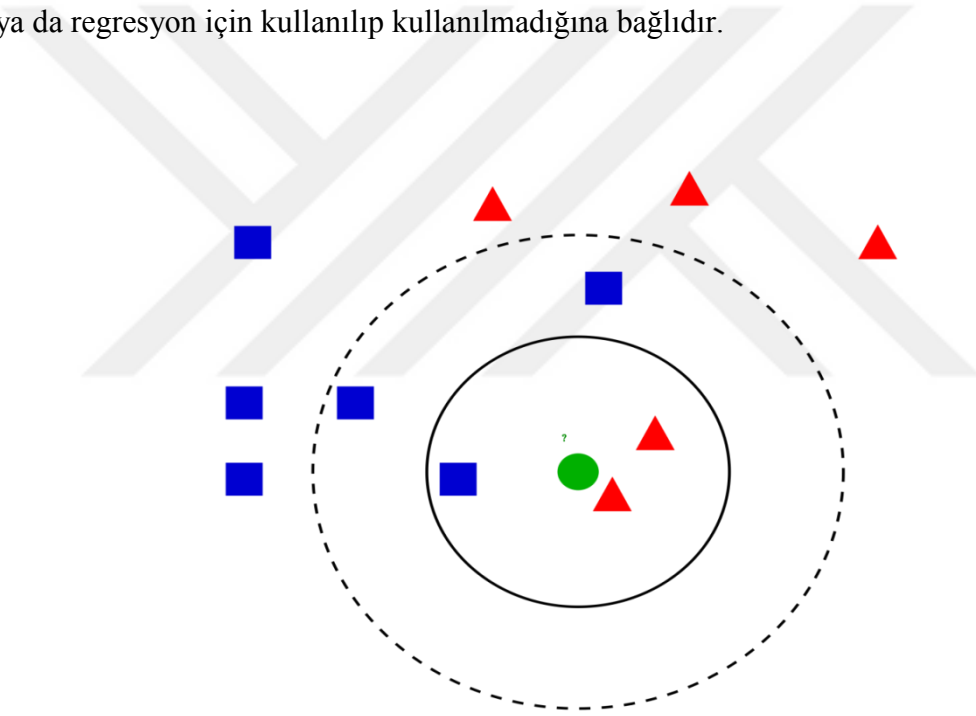
Geleneksel sınıflandırma yaklaşımları, verilerin yüksek boyutluluğu nedeniyle doğrudan çalışırken kötü performans gösterir, ancak destek vektör makineleri çok yüksek boyutlu gösterimlerin tuzaklarını önleyebilir(J.P.Lewis,2004). Metin sınıflandırma için açıklanan tekniklere çok benzer bir yaklaşım, görüntü sınıflandırma görevi için de kullanılabilir ve bu durumda olduğu gibi, doğrusal zor kenar makineleri sıklıkla iyi bir şekilde genelleştirilebilir(Nello Cristianini and John Shawe-Taylor,2000).

Destek Vektörü Makinelerinin test edildiği ilk gerçek dünya görevi, elle yazılmış karakter tanıma sorunu olmuştur. Ayrıca, çok sınıflı SVM'ler bu veriler üzerinde test edilmiştir. Yalnızca SVM'leri diğer sınıflandırıcılarla karşılaştırmak değil, aynı zamanda kendi aralarındaki farklı SVM'leri karşılaştırmak da ilginçtir (M. O. Stitson and J. A. E. Weston,1996). Yaklaşık aynı performansa sahip oldukları ve ayrıca

seçilen çekirdeklerinden bağımsız olarak destek vektörlerinin çoğunu paylaştığı ortaya çıkmıştır. SVM'nin bu sistemlerin yanı sıra detaylı bir ön bilgi dahil etmeden de yapabileceği gerçeği kesinlikle dikkat çekicidir (Osuna E., Freund R., and Girosi F.,1997).

3.6. K-NN Yakın Komşu Algoritması

En yakın komşu algoritması (K- nn), örüntü tanımada sınıflandırma ve regresyon için kullanılan parametrik olmayan bir yöntemdir (Altman, 1992). Giriş, özellik alanındaki en yakın eğitim örneklerinden oluşmaktadır, çıktı k-NN'nin sınıflandırma ya da regresyon için kullanılıp kullanılmadığına bağlıdır.

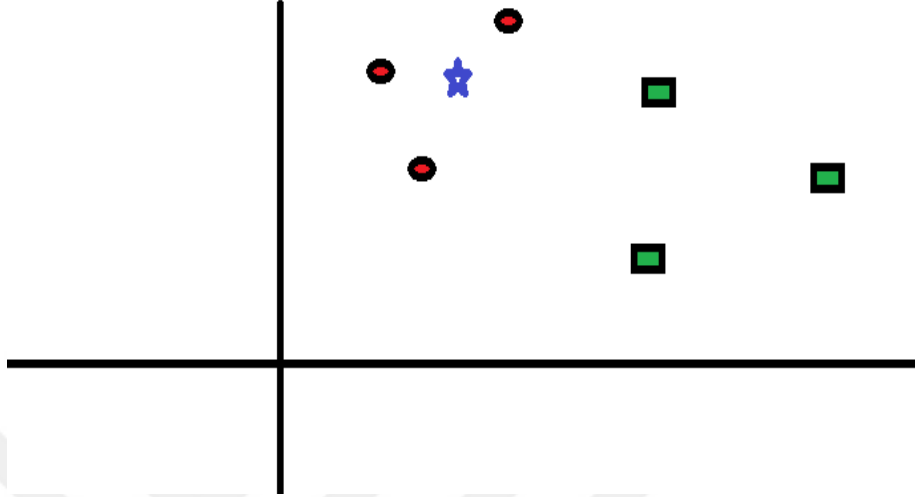


Şekil 3.16. KNN en yakın komşu algoritması

KNN hem sınıflandırma hem de regresyon tahmini problemleri için kullanılabilir. Ancak, endüstrideki sınıflandırma problemlerinde daha yaygın olarak kullanılmaktadır. Herhangi bir tekniği değerlendirmek için genellikle 3 önemli konuya bakılmaktadır: 1. Çıktıyı yorumlama kolaylığı 2. Hesaplama süresi 3. Tahmini Güç KNN'yi ölçeğe yerleştirmesine.

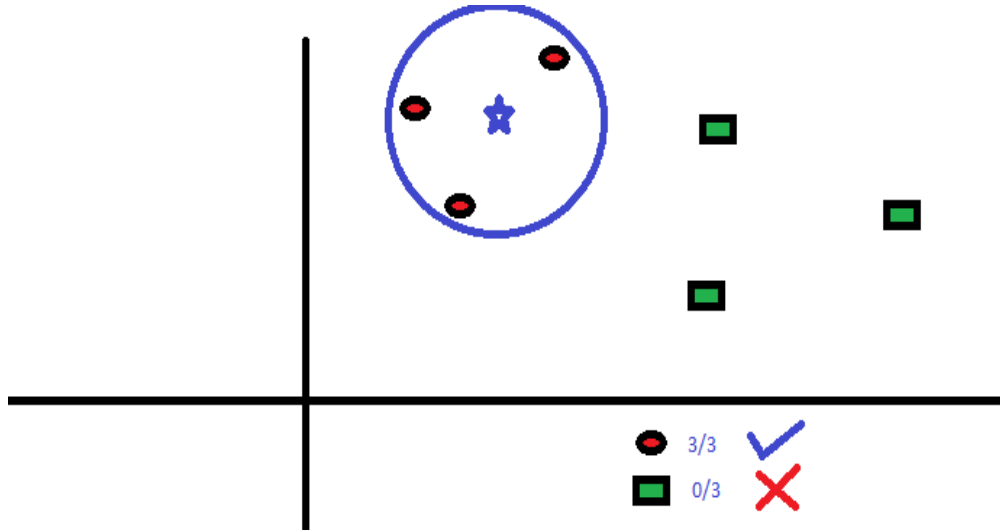
3.6.1. KNN algoritması nasıl çalışır

Bu algoritmayı anlamak için basit bir örneği ele alalım. Aşağıdaki kırmızı daireler (RC) ve yeşil kareler (GS) yayılmasıdır:



Şekil 3.11. (RC)v(GS)yayılması

Mavi yıldızın (BS) sınıfını bulmak gerekmektedir. BS, RC veya GS olabilir başka olamaz. Örneğin $k = 3$ için yeni bir eleman sınıflandırılmak istendiğinde, BS ile uçağa sadece üç veri noktası dahil edecek kadar büyük bir merkez yapmak gerekir. Daha detaylı ayrıntı şekilde gösterilmektedir.



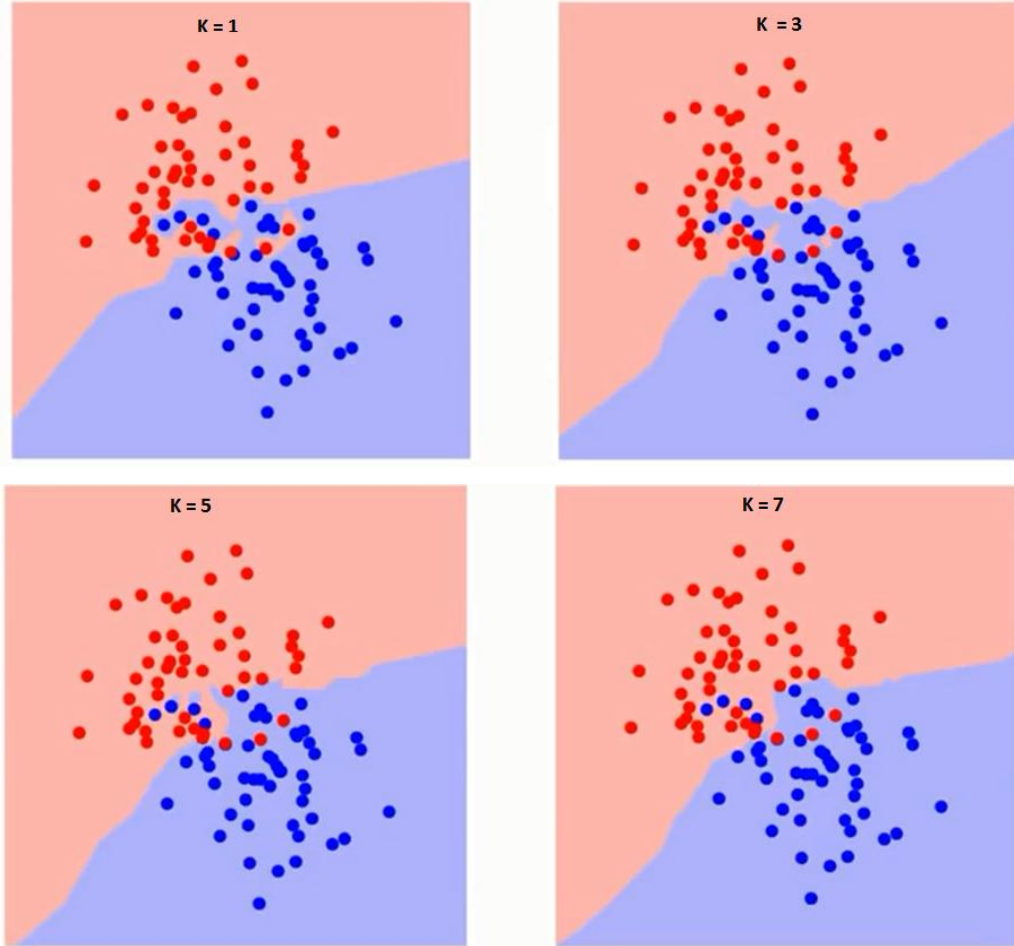
Şekil 3.17. (BS),(RC)ve(GS) sonuçları

BS'ye en yakın üç nokta RC noktasıdır. Bu nedenle, iyi bir güven seviyesiyle, BS'nin RC sınıfına ait olması gerektiği söylenebilir. Burada, en yakın komşunun üç oyu

da RC'ye gider bu yüzden seçim çok belirgindir. K parametresinin seçimi bu algoritmada çok önemlidir. Daha sonra, en iyi sonucu elde etmek için göz önünde bulundurulması gereken faktörlerin ne olduğu görülecektir.

3.6.2. K faktörünü nasıl seçeriz?

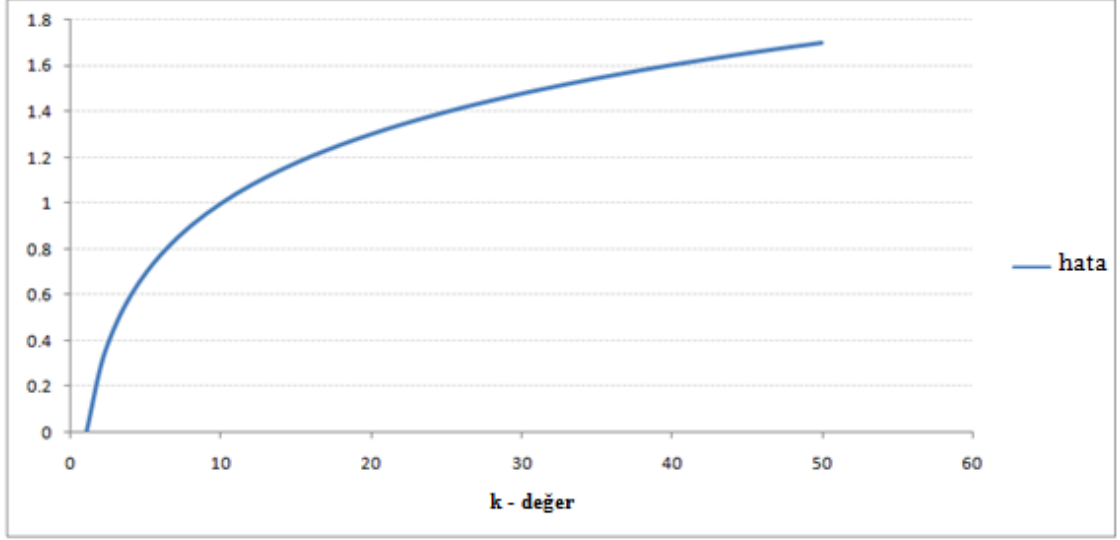
Öncelikle, K'nin algoritmada tam olarak neyi etkilediğini anlaşılmalıdır. Son örneğe bakıldığında, 6 eğitim gözleminin tamamı sabit kalırsa verilen bir K değeri ile her sınıfın sınırlarını yapılabilir. Bu sınırlar RC'yi GS'den ayıracaktır. Aynı şekilde, “K” değerinin sınıf sınırları üzerindeki etkisi görülecektir. Şekil 3.13.' ki iki sınıfı K'nın farklı değerleri ile ayıran farklı sınırlar verilmiştir.



Şekil 3.18. iki sınıfı K'nın farklı değerleri ile ayıran farklı sınırlar verilmiştir

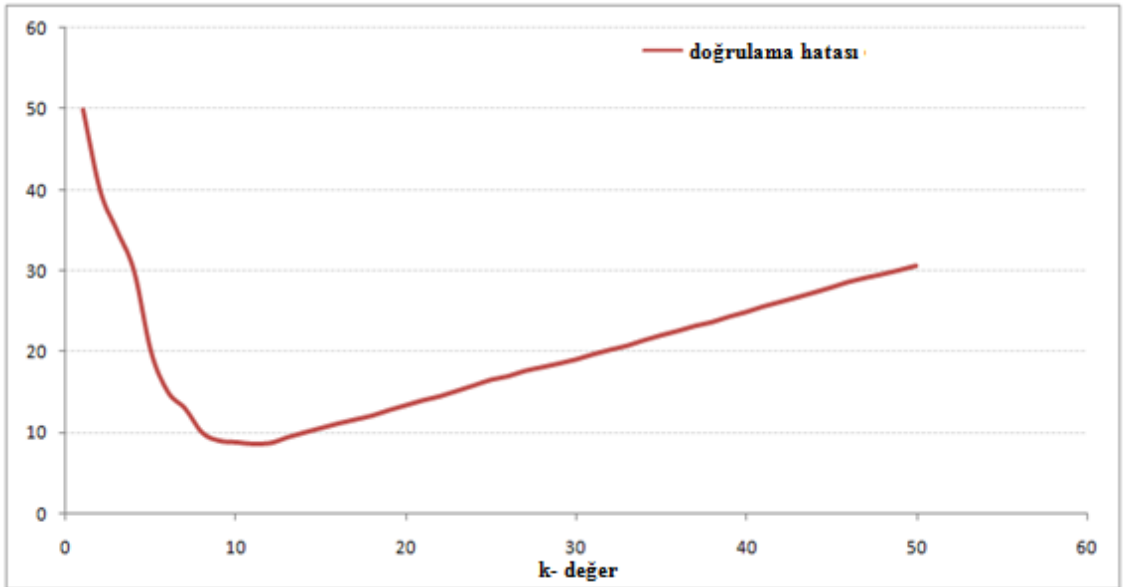
Dikkatli bir şekilde bakıldığında, sınırın K değeri ile daha pürüzsüz hale geldiğini görülür. K değeri arttıkça, toplam çoğunluğa bağlı olarak tamamen mavi veya

tamamen kırmızı olur. Hata oranı ve doğrulama hata oranı, farklı K değerlerine erişilmesi gereken iki parametredir. Şekil 3.19.'ki değişen K değerinde olan hata oranı eğrisi şöyledir:



Şekil 3.19. Değişen K değerinde olan hata oranı eğrisi göstermektedir.

Gördüğü gibi, $K = 1$ 'deki hata oranı, numunesi için her zaman sıfırdır. Bunun nedeni, herhangi bir antrenman veri noktasına en yakın noktanın kendisinin olmasıdır. Bu nedenle, tahmin $K = 1$ ile her zaman doğrudur. Doğrulama hatası eğrisi benzer olsaydı, K seçimi 1 olurdu. Aşağıda, K değeri değişen doğrulama hatası eğrisi gösterilir:



Şekil 3.20. K değeri değişen doğrulama hatası eğrisi

3.6.3. Ağırlıklı en yakın komşu sınıflandırıcı

En yakın komşulara ağırlık olarak en yakın komşu sınıflandırıcı $1/k$ ve diğerleri için ağırlık olarak 0 atanabilir. Bu durum, en yakın komşu sınıflandırıcılar için ağırlıklandırılabilir. Yeni ağırlık en yakın komşuya verilmektedir. w_{ni} , ile $\sum_{i=1}^n w_{ni} = 1$ Ağırlıklı en yakın komşu sınıflandırıcılarının güçlü tutarlılığına dair benzer bir sonuç da bulunmaktadır (Stone, 1977). $C_n^{w_{nn}}$ Ağırlıklı en yakın sınıflandırıcı ağırlıkları ile belirtilir $\{w_{ni}\}_{i=1}^n$. Sınıf dağılımlarındaki düzenlilik şartlara bağlı olarak aşırı risk aşağıdaki asimptotik genişlemeye sahip olmaktadır (Samworth, 2012).

$$R_R(C_n^{w_{nn}}) - R_R(C^{Bayes}) = (B_1 s_n^2 + B_2 t_n^2) \{1 + o(1)\}, \text{ sabitler için } B_1 \text{ ve } B_2$$

$$s_n^2 \text{ ve } t_n^2 = n^{-2/d} - \sum_{i=1}^n w_{ni} \{i^{1+2/d} - (i-1)^{1+2/d}\} \quad (3.26)$$

Optimum ağırlıklandırma şeması $\{w_{ni}^*\}_{i=1}^n$, Yukarıdaki ekranda iki terimi dengeleyen, aşağıdaki şekilde gösterilmiştir.

$$k^* = \lfloor B_n \rfloor, \quad (3.27)$$

$$w_{ni}^* = \frac{1}{k^*} \left[1 + \frac{d}{2} \{i^{1+2/d} - (i-1)^{1+2/d}\} \right] \text{ için } i = 1, 2, \dots, k^* \text{ ve} \quad (3.28)$$

$$w_{ni}^* = 0 \text{ için } i = k^* + 1, \dots, n. \quad (3.29)$$

Optimal ağırlıklarla aşırı riskin asimptotik genişlemesinde baskın terim $\mathcal{O}(n)$ 'dir. Torbalı en yakın komşu sınıflandırıcı kullanıldığı zaman benzer sonuçlarda geçerlidir.

3.7. Konuşma Tanıma Uygulamalarında Sınıflandırma Yöntemleri

Literatürde ses ile ilgili uygulamalar çok fazla bulunmaktadır. Ses sinyalleriyle ilgili olarak yapılan çalışmaların sınıflandırılması iki farklı kritere göre yapılmıştır. Konuşma tanıma sistemlerinin sınıflandırılması aşağıdaki gibidir; (Hopfield, 1982; Elmas, 2003; Gökberk ve ark., 2003).

- Kelime bağlı konuşma tanıma.
- Fonem bağlı konuşma tanıma.

3.7.1. Konuşulan konuşma türleri

Konuşma tanıma tekniği, hangi tür ifadeleri tanıma konusunda yetkin olduklarına ve aşağıda gösterildiğine bağlı olarak çeşitli gruplara ayrılabilir.

3.7.1.1. İzole konuşma

İzole edilmiş kelime tanıyıcıları, örnek pencerenin her iki tarafında bir ses sinyali eksikliği gibi sessizliğe sahip olmak için her bir ifadeye ihtiyaç duyar. Belirli bir zamanda tek kelime veya sözler alır. Bu teknik, kullanıcının duraklamalar sırasındaki işlemini alışkanlık haline getiren sözler arasında beklemesi gereken “dinleme/dinleme-dinleme” durumlarını içerir. Daha iyi izole ifade olarak adlandırılabilir.

3.7.1.2. Bağlı kelimeler

Bağlantılı kelime, konuşma akışını zahmetsizce sağlamak için ifadeler arasında minimum bir duraklamaya ihtiyaç duyar. İzole edilen kelimelerle hemen hemen aynıdır.

3.7.1.3. Sürekli konuşma

Sürekli konuşma tanıyıcıları, kullanıcıların içeriği kararlaştırırken kullanıcıların daha fazla veya daha az doğal konuşmalarına izin verir. Temel olarak, bilgisayar diktesini temsil eder. Devamlı konuşma becerisine sahip tanıyıcılar, söyleme sınırlarına karar vermek için benzersiz teknikler kullandıkları için üretilmesi daha zordur.

3.7.1.4. Spontane konuşma

Temel bir seviyede, eğitilmiş bir konuşma değil, doğal bir sondaj olan konuşma kabul edilebilir. Spontan konuşma kapasitesine sahip bir ASR sistemi, hafif öbekler de

dahil olmak üzere “ums” ve “ahs” gibi birlikte çalışan kelimeler gibi bir dizi doğal konuşma özelliğini ele alma yetkinliğine sahip olmalıdır.

3.7.2. Uygulama çerçevesi

İnsan sesi uzun yıllardır kapsamlı bir şekilde araştırılmış olsa da, hem oluşumunda hem de doğasında hala çok bilinemeyeni olan çok karmaşık bir olgudur. Bu nedenle, bu çalışmada konuşma tanıma uygulaması belirli bir çerçeve ile sınırlıdır. Bu bağlamda konuşmacıya bağımlı olma, kelime tabanlı olma, izole edilmiş kelimeler kullanma, metine dayalı olma kriterlerine sahip bir Konuşma Tanıma sisteminin tasarımı üzerine odaklanılmıştır (Elmas, 2003; Gökberk ve ark., 2003; Oral ve Çelik, 2003).

3.7.3. Performans değeri ölçütleri

Formül 3.31.'de duyarlılık, kesinlik, özgünlük ve hasssiyet'in performans değerlerinin nasıl hesaplandığı verilmiştir.

Burada da Sınıflandırma Oranı veya Doğruluk ilişkisi verilir.

$$\text{Doğruluk} = \frac{\text{Gerçek Olumlu}(TP)+\text{Gerçek olumlu}}{\text{Toplam Örnekler}} \quad (3.30)$$

Duyarlılık, toplam olarak sınıflandırılan pozitif örneklerin toplam sayısının, toplam pozitif örnek sayısına bölünmesi olarak tanımlanabilir. Yüksek Geri Çağırma, sınıfın doğru bir şekilde tanındığını gösterir (az sayıda FN). Duyarlılık eşittir Hassasiyet (Recall) aynı formüle kullanmaktadır. Duyarlılık ilişki tarafından verilir:

$$\text{Duyarlılık}=(\text{Gerçek Olumlu})/(\text{Gerçek Olumlu}+ \text{ Yanlış Negatif}) \quad (3.31)$$

Kesinlik, yanlış pozitiflerin (TN) maliyetlerinin yüksek olması durumunda yardımcı olur. Aşağıdaki formül göstermektedir.

$$\text{Kesinlik} = \text{Gerçek Olumlu} \frac{(TP)}{\text{Gerçek Olumlu}(TP)+\text{Yanlış Olumlu}} \quad (3.32)$$

Özgünlük, negatif sonuçların bölünmesiyle oluşan doğru negatif tahminlerin sayısı olarak hesaplanır. Aynı zamanda gerçek negatif oran (TNR) olarak da adlandırılır. Aşağıdaki formül göstermektedir.

$$\text{Özgünlük} = \frac{\text{Gerçek Olumsuz}}{\text{Gerçek Olumsuz} + \text{Yanlış}} \quad (3.33)$$



4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

4.1. Önerilen Metod

Bu tezde, parçacık sürüsü optimizasyonu ve özellikler çıkarımı için doğrusal öngörüm kodlamasına (Lpc) dayanan bir konuşma tanıma sistemi kullanılmıştır, üç sınıflandırma algoritması ile özellikle seçim amacı için PSO algoritması kullanılmıştır. Her biri beş kelimeye sahip olan on iki kişiden toplanan veri setinin (back, left, right, stop ve go) toplam eğitim seti sayısı 60 kayıttır, her dosyada bir saniye uzunluğundaki ses 8000Hz'de örneklendirilmiştir, tüm dosyalar arka plan gürültüsü olmaksızın izole edilmiş bir ortamda kaydedilmiştir. Test seti, aynı uzunluk ve örnekleme oranına sahip, farklı bir günde kaydedilen 40 kayıt şeklindedir.

PSO, bu çalışmada Şekil 4.1'de gösterildiği gibi bir özellik numarası seçim yöntemi olarak uygulanır. Öncelikle program, eğitim ve test verilerini yükler ve ardından her ses dosyasından LPC özelliklerini çıkarır. LPC özelliklerinin sayısı 200, PSO, her bir ses dosyasının 200 LPC özelliklerinden uygun bir özellik numarası(kümesi) belirler, daha sonra bu çalışmada kullanılan SVM, KNN ve YSA gibi seçilen sınıflandırıcıyı eğitir, daha sonra seçilen her özellik numarasının doğruluğunu hesaplar ve seçilen bu özelliklerle test setinin en iyi doğruluğu seçilir. Doğruluk PSO iyileştirilmesi değilse başka özellikler uzunluğu çalışılacaktır.

Kullanılan optimize edilmiş Uygunluk işlevi:

$$Uygunluk = 0.2 \frac{Seçilen\ Özelliklerin\ Sayısı}{Toplam\ Özelliklerin\ Sayısı} - 0.8(Test\ Doğruluğu) \quad (4.1)$$

Uygunluk fonksiyonu, çözümünüzün ne kadar iyi olduğunu ölçen bir fonksiyon olmalıdır. Özellikle, üretilen çözümler ne olursa olsun onu idare edebilmelidir ve bunları geliştirmek için doğru yolu göstermelidir.

Örneğin, cevap doğru olmadığı sürece sıfır olan bir uygunluk işlevi iyi değildir, çünkü çözümün doğru cevaba ne kadar yakın olduğu hakkında bir fikir edinmenize yardımcı olmaz. Ayrıca, çözümler iyileştikçe artan, ancak en iyi çözümü tanımlamayan

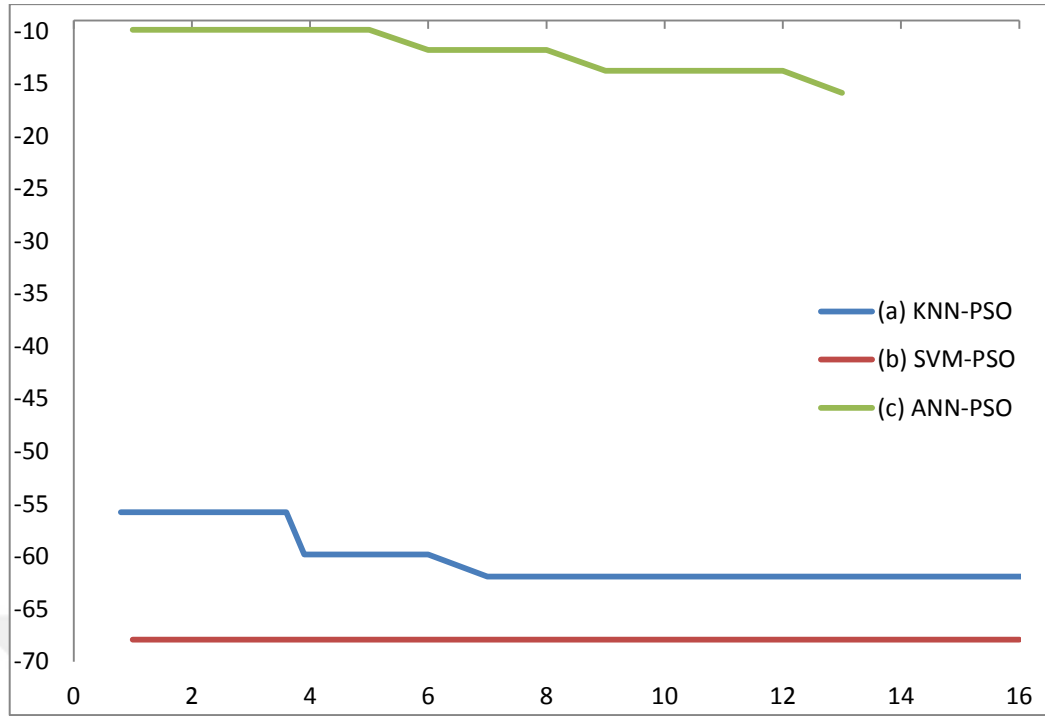
bir fitness işlevi de o kadar iyi değildir, çünkü popülasyon belirli bir noktaya kadar gelişecek ve sonra sıkışacaktır (Wong ve ark., 2013).

Uygunluk fonksiyonu bu şekilde tanımlanabilir:

Bu özellik, bir ürünün veya sistemin, belirtilen koşullar altında kullanıldığında belirtilen ihtiyaçları karşılayan işlevleri sağlama derecesini gösterir. Burda'da fitness in amacı az nitelik ile yüksek doğruluk elde etmek.



Şekil 4.1. Sistemin Akış Şeması



Şekil 4.2. (a) KNN-PSO yöntemi için uygunluk fonksiyonunun en iyi çözüm eğrisi, (b) SVM-PSO yöntemi için uygunluk fonksiyonunun en iyi çözüm eğrisi, (c) ANN-PSO yöntemi için uygunluk fonksiyonunun en iyi çözüm eğrisi .

Çizelge 4.1. Test Performans Değeri

	Kesinlik	Özgünlük	Hassasiyet
KNN	62.50%	62.50%	29.41%
KNN+PSO	62.50%	81.25%	40%
SVM	75%	93.75%	50%
SVM+PSO	62.50%	81.25%	66.67%
ANN	50%	78.13%	20%
ANN+PSO	25%	40.6%	9.5%

Çizelge 4.2. Eğitim Performans Değeri

	Kesinlik	Özgünlük	Hassasiyet
KNN	100%	100%	100%
KNN+PSO	100%	100%	100%
SVM	100%	100%	100%
SVM+PSO	100%	100%	100%
ANN	66.67%	81.25%	47%
ANN+PSO	83.3%	68.75%	40%

Çizelge 4.3. SVM'nin Konfüzyon Matrisi

		GERÇEK					DOĞRU	YANLIŞ	TOPLAM
		BACK	GO	LEFT	RIGHT	STOP			
TAHMİN	BACK	5	0	1	2	0	5	3	8
	GO	0	7	0	1	0	7	1	8
	LEFT	2	0	6	0	0	6	2	8
	RIGHT	1	0	0	7	0	7	1	8
	STOP	0	0	0	1	7	7	1	8
	TOPLAM						32	8	40

Çizelge 4.3 e baktığımızda toplam 8 örneğimizden BACK'de 5 doğru 3 yanlış ,GO'da 7 doğru 1 yanlış ,LEFT 6 doğru 2 yanlış ,RIGHT 7 doğru 1 yanlış ve STOP 7 doğru 1 yanlış, toplam 32 doğru 8 yanlış bulunmaktadır .

Çizelge 4.4. SVM+PSO'nun Konfüzyon Matrisi

		GERÇEK					DOĞRU	YANLIŞ	TOPLAM
		BACK	GO	LEFT	RIGHT	STOP			
TAHMİN	BACK	7	0	0	1	0	7	1	8
	GO	0	7	0	1	0	7	1	8
	LEFT	0	0	8	0	0	8	0	8
	RIGHT	1	0	0	7	0	7	1	8
	STOP	0	0	0	1	7	7	1	8
	TOPLAM						36	4	40

Çizelge 4.4 e baktığımızda toplam 8 örneğimizden BACK'de 7 doğru 1 yanlış ,GO'da 7 doğru 1 yanlış ,LEFT 8 doğru ,RIGHT 7 doğru 1 yanlış ve STOP 7 doğru 1 yanlış, toplam 36 doğru 4 yanlış bulunmaktadır.

Çizelge 4.5. KNN'nin Konfüzyon Matrisi

		GERÇEK					DOĞRU	YANLIŞ	TOPLAM
		BACK	GO	LEFT	RIGHT	STOP			
TAHMİN	BACK	5	0	3	0	0	5	3	8
	GO	1	5	2	0	0	5	3	8
	LEFT	1	0	7	0	0	7	1	8
	RIGHT	2	0	2	4	0	4	4	8
	STOP	0	0	3	1	4	4	4	8
	TOPLAM						25	15	40

Çizelge 4.5 e baktığımızda toplam 8 örneğimizden BACK'de 5 doğru 3 yanlış ,GO'da 5 doğru 3 yanlış ,LEFT 7 doğru 1 yanlış ,RIGHT 4 doğru 4 yanlış ve STOP 4 doğru 4 yanlış , toplam 25 doğru ve 15 yanlış bulunmaktadır.

Çizelge 4.6. KNN+PSO'nun Konfüzyon Matrisi

		GERÇEK					DOĞRU	YANLIŞ	TOPLAM
		BACK	GO	LEFT	RIGHT	STOP			
TAHMİN	BACK	5	1	2	0	0	5	3	8
	GO	1	5	2	0	0	5	3	8
	LEFT	0	0	8	0	0	8	0	8
	RIGHT	0	0	0	7	1	7	1	8
	STOP	0	0	1	1	6	6	2	8
	TOPLAM						31	9	40

Çizelge 4.6 e baktığımızda toplam 8 örneğimizden BACK'de 5 doğru 3 yanlış ,GO'da 5 doğru 3 yanlış ,LEFT 8 doğru ,RIGHT 7 doğru 1 yanlış ve STOP 6 doğru 2 yanlış, toplam 31 doğru ve 9 yanlış bulunmaktadır.

4.2. Araştırma Sonuçları

12 kişiden farklı yaş grubundaki konuşmacının her 5 kelimeyi birer kez tekrarlama sonucu her bir kelime için 60 adet eğitim verisi ve 40 adet test verisi elde edilmiştir. Geliştirilen uygulamada kodlayıcı olarak, ses özelliklerini de dikkate alan LPC kodlayıcı kullanılmaktadır. KNN, YSA ve SVM'nin sınıflandırmada önemli bir

noktası olan öğrenme aşamasında sezgisel bir algoritma olan Parçacık Sürü Optimizasyonu (PSO) kullanılarak sınıflandırmanın başarısı daha da yukarı çekilmiştir. PSO Rastgele üretilen başlangıç pozisyonu ve hız ve yönü ile başlangıç sürüsü oluşturur, kişiye dayalı eğitim ve testler için uygun bir özellik numarası belirler, tüm parçacıkların uygunluk değerleri hesaplanır, her bir parçacık için en iyi yaklaşım bulunur, eğer sonuca yeterince yaklaşıldı ise yani doğruluk gerçekleştiyse seçilen özelliklerin en iyisidir, yoksa bu adımlar tekrar yapılır. Bu işlem sonucunda her bir Konfüzyona karşılık gelen doğruluk oranları Çizelge 4.7’de ve Çizelge 4.8’de verilmiştir.

Çizelge 4.7. Kullandığımız yöntemlerin test doğruluk oranı

Yöntem	Test Seti örnekleri	Doğruluk
SVM	40 örnekler	80%
PSO+SVM	40 örnekler	90%
KNN	40 örnekler	62,5%
KNN+PSO	40 örnekler	77,5%
YSA	40 örnekler	10%
YSA+PSO	40 örnekler	20%

Çizelge 4.8. Kullandığımız yöntemlerin eğitim doğruluk oranı

Yöntem	Test Seti örnekleri	Doğruluk
SVM	40 örnekler	100%
PSO+SVM	40 örnekler	100%
KNN	40 örnekler	100%
KNN+PSO	40 örnekler	100%
YSA	40 örnekler	10%
YSA+PSO	40 örnekler	40%

Çizelge 4.7.’ e bakıldığında bu tez’de konuşma tanıma sisteminde sınıflandırmada sadece En Yakın Komşuluk Algoritması (KNN) kullandığımız zaman doğruluk oranı %60,00 iken beraber parçacık sürü optimizasyon algoritması kullandığımızda doğruluk oranı %77,5’e çıkmıştır.

Ayrıca yalnızca SVM kullandığımız zaman doğruluk oranı %80,00dir, PSO ve SVM beraber kullanıldığı zaman parametrelerinin en yüksek başarı oranını %90 yakaladığı görülmektedir ve en kötü sonuçları YSA elde etmiştir, YSA sadece %10 test

setinin bir doğruluđuna sahiptir PSO ile beraber kullanıldıđı zaman %20 doğruluđuna ıkartılmıřtır.

izelge 4.8.' e bakıldıđında kullandıđımız eđitim yntemlerin dođruluk oranı, KNN,KNN+PSO,SVM ve SVM+PSO dođruluk oranı %100'ıkmaktadır. Sadece YSA %10 eđitim setinin bir dođruluđuna sahiptir PSO ile beraber kullanıldıđı zaman %40 dođruluđuna ıkartılmıřtır.



5. SONUÇLAR VE ÖNERİLER

5.1 Sonuçlar

Bu tezde, Parçacık Sürüsü Optimizasyonuna ve özellik çıkarımı için Doğrusal tahmin kodlamasına (LPC) dayanan bir konuşma tanıma sistemi önerilmiştir, üç sınıflandırma algoritmasıyla özellik seçimi amacıyla PSO algoritması kullanılmıştır. SVM, KNN ve YSA algoritması kullanılmıştır ve LPC özellikleri konusunda eğitilmiştir.

SVM-PSO yöntemi, test setinde% 90 doğrulukla daha yüksek sonuçlar elde etmiştir, SVM ise sadece% 80 doğruluk elde etmiştir. KNN-PSO metodu %77,5 doğruluğa, KNN ise sadece% 62,5'e ulaşmaktadır. En kötü sonuçları YSA elde etmiştir, YSA sadece %10 test setinin bir doğruluğuna sahiptir PSO ile beraber kullanıldığı zaman %20 doğruluğuna çıkartılmıştır.

Çizelge 4.8.' e bakıldığında kullandığımız eğitim yöntemlerin doğruluk oranı, KNN, KNN+PSO, SVM ve SVM+PSO doğruluk oranı %100' çıkmaktadır. Sadece YSA %10 eğitim setinin bir doğruluğuna sahiptir PSO ile beraber kullanıldığı zaman %40 doğruluğuna çıkartılmıştır.

Sonuç olarak bu çalışmada, mevcut tekniklerden yararlanılarak ses tanıma oluşturulmuştur ve elde edilen bu sistem ile bu alanda çalışma yapacak isteyen kişilere bu tekniklerin uygulama içinde kullanımı ve ilişkilendirilmesi hakkında örnek teşkil etmesi amaçlanmıştır.

Tasarlanan bu sistemin başarısının artırılması ile fiziksel engelli olan kişilerin bilgisayar kullanımının sağlanması, araç sürücülerinin belli işlevleri sesli komutlarla gerçekleştirmeleri ve üretim sahasından veri toplama amacıyla kullanımının sağlanması amaçlanmaktadır.

5.2 Öneriler

Daha çok ses veri tabanlarını seçerek sistemin doğruluk oranını test edilmesi ve daha iyi başarı oranlarının elde etmek için araştırılmalıdır.

Eğer diğer sınıflandırma algoritmaları ve öğrenme algoritmaları uygulanırsa, sistem performansını daha yüksek seviyelere getirebilir, C++ kullanılabilir veya makine dili gibi farklı programlama dilleri kullanılabilir.

Bu gibi yöntemlerin geliştirilmesiyle video kameralardan alınan görüntü ile, parmak iziyle, doğal konuşma ile, göz bebeğinin yapısıyla bile kimlik tanıyabilen güvenlik sistemleri geliştirilebilecektir.

KAYNAKLAR

- Ahad, A., Fayyaz, A. ve Mehmood, T., 2002, Speech recognition using multilayer perceptron, *IEEE Students Conference, ISCON'02. Proceedings.*, 103-109.
- Altman, N. S. J. T. A. S., 1992, An introduction to kernel and nearest-neighbor nonparametric regression, 46 (3), 175-185.
- Babu, C. G., Kumar, R. H. ve Vanathi, P., 2012, Performance analysis of hybrid robust automatic speech recognition system, *2012 IEEE International Conference on Signal Processing, Computing and Control*, 1-4.
- Benesty, J., Sondh, M. ve Huang, Y. J. N. S., 2008, Springer handbook of speech recognition, 1176.
- Bolia, R. S., Nelson, W. T., Ericson, M. A. ve Simpson, B. D. J. T. J. o. t. A. S. o. A., 2000, A speech corpus for multitalker communications research, 107 (2), 1065-1066.
- Borden, G. J. S. s. p. B., Maryland, Williams ve Wilkins, 1984, Harris.
- Bronkhorst, A. W. J. A. A. u. w. A., 2000, The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions, 86 (1), 117-128.
- Cai, M. ve Liu, J., 2016, Maxout neurons for deep convolutional and LSTM neural networks in speech recognition, *Speech Communication*, 77, 53-64.
- Campbell, W. M., Campbell, J. P., Reynolds, D. A., Singer, E., Torres-Carrasquillo, P. A. J. C. S. ve Language, 2006, Support vector machines for speaker and language recognition, 20 (2-3), 210-229.
- Dede, G. ve Sazlı, M. H., 2010, Speech recognition with artificial neural networks, *Digital Signal Processing*, 20 (3), 763-768.
- Denes, P. B. ve Pinson, E. N. J. I. B., Maryland: Waverly Press, Inc, 1963, The Speech Chain: The Physics and Biology of Spoken Language. Bell Telephone Laboratories.
- Elmas, Ç. J. M., Eğitim, Uygulama, Seçkin Yayıncılık, Ankara, 2003, Yapay Sinir Ağları Kuram.
- Ergezer, H., Dikmen, M. ve Özdemir, E. J. P., 2003, Yapay sinir ağları ve tanıma sistemleri, 2 (6), 14-17.
- Farley, B. ve Clark, W. J. T. o. t. I. P. G. o. I. T., 1954, Simulation of self-organizing systems by digital computer, 4 (4), 76-84.

- Farooq, O. ve Datta, S. J. E. I., 2001, Robust features for speech recognition based on admissible wavelet packets, 37 (25), 1554-1556.
- Fukushima, K. J. B. c., 1975, Cognitron: A self-organizing multilayered neural network, 20 (3-4), 121-136.
- Gökberk, B., İrfanoğlu, M. O., Doğu, H., Akarun, L. ve Alpaydın, E., 2003, Yüz Tanıma İçin Eniyi Gabor Parametrelerinin Seçimi, SIU.
- Grieder, W. ve Kinsner, W., 1994, Speech segmentation by variance fractal dimension, *Canadian Conference on Electrical and Computer Engineering*, 481-485.
- Haykin, S. J. N. n. a. c. f., 1999a, Multilayer perceptrons, 2, 156-255.
- Haykin, S. J. U. S. R., New Jersey, 1999b, Neural Networks, a comprehensive foundation, Prentice-Hall Inc, 7458, 161-175.
- Hermansky, H. J. t. J. o. t. A. S. o. A., 1990, Perceptual linear predictive (PLP) analysis of speech, 87 (4), 1738-1752.
- Hopfield, J. J. J. P. o. t. n. a. o. s., 1982, Neural networks and physical systems with emergent collective computational abilities, 79 (8), 2554-2558.
- Jiang, Z., Huang, H., Yang, S., Lu, S. ve Hao, Z., 2009, Acoustic feature comparison of MFCC and CZT-based cepstrum for speech recognition, *2009 Fifth International Conference on Natural Computation*, 55-59.
- Kecman, V., 2001, Learning and soft computing: support vector machines, neural networks, and fuzzy logic models, MIT press, p.
- Kumar, A., et al. (2016). "Particle Swarm Optimization: A Study of Variants and Their Applications." 135(5): 24-30.
- Kraetzer, C., Oermann, A., Dittmann, J. ve Lang, A., 2007, Digital audio forensics: a first practical evaluation on microphone and environment classification, *Proceedings of the 9th workshop on Multimedia & security*, 63-74.
- Leechor, P., Pornpanomchai, C. ve Sukklay, P., 2010, Operation of a radio-controlled car by voice commands, *2010 2nd International Conference on Mechanical and Electronics Engineering*, V1-14-V11-17.
- Luo, Y., Liu, Y., Zhang, Y. ve Yue, C., 2018, Speech bottleneck feature extraction method based on overlapping group lasso sparse deep neural network, *Speech Communication*, 99, 56-61.
- Olson, H. F., 1972, Modern sound reproduction, Van Nostrand Reinhold, p.
- Oral, M. ve Çelik, U., 2003, MOTORLU ARAÇ PLAKA GÖRÜNTÜLERİNDEN KARAKTER AYRIŞTIRMA VE TANIMA.

- Samworth, R. J. J. T. A. o. S., 2012, Optimal weighted nearest neighbour classifiers, 40 (5), 2733-2763.
- Slaney, M. J. I. R. C., Tech. Rep, 1998, Auditory toolbox, 10 (1998).
- Stone, C. J. J. T. a. o. s., 1977, Consistent nonparametric regression, 595-620.
- Tunçkanat, M., Kurban, R. ve Sagoroglu, S., 2003, Voice recognition based on neural networks, *International Conference On Signal Processing, ICSP*.
- Wong, C., Guo, Z. X. ve Leung, S., 2013, Optimizing decision making in the apparel supply chain using artificial intelligence (AI): from production to retail, Elsevier, p.
- Wong, E. ve Sridharan, S., 2001, Comparison of linear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification, *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing. ISIMP 2001 (IEEE Cat. No. 01EX489)*, 95-98.
- Zhang, L., Xu, X., Chen, H., Chen, J. ve Ye, Z., 2018, Supervised single-channel speech dereverberation and denoising using a two-stage model based sparse representation, *Speech Communication*, 97, 1-8.
- Zhi, Q., Kaynak, M., Sengupta, K., Cheok, A. ve Ko, C., 2001, A study of the modeling aspects in bimodal speech recognition, *Proc. 2001 IEEE International Conference on Multimedia and Expo (ICME2001)*.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Arzo Mahmood
Uyruğu : IRAK
Doğum Yeri ve Tarihi : 06/08/1992 IRAK/KERKÜK
Telefon : 05378165077
Faks :
e-mail : Janamohammed22@gmail.com

EĞİTİM

Derece	Adı, İlçe, İl	Bitirme Yılı
Lise	: Ayqada Kiz Lisesi Kerkük	2010
Üniversite	: AL Qalem Üniversitesi Kerkük	2014
Yüksek Lisans	: Konya Teknik Üniversitesi konya	
Doktora	:	

UZMANLIK ALANI

YABANCI DİLLER

Arabça ,İngilizce

BELİRTMEK İSTEĞİNİZ DİĞER ÖZELLİKLER

YAYINLAR

Konu: Comparative Study on Automatic Speech Recognition /
Yer: Konya Technical University / ICENTE'18