İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

DERİNLİKLİ ÖĞRENME KULLANILARAK KONUŞMADAN UYKULULUK/UYKUSUZLUK TESPİTİ

YÜKSEK LİSANS TEZİ

Canberk HACIOĞLU

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Telekomünikasyon Mühendisliği Programı

MAYIS 2014

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

DERİNLİKLİ ÖĞRENME KULLANILARAK KONUŞMADAN UYKULULUK/UYKUSUZLUK TESPİTİ

YÜKSEK LİSANS TEZİ

Canberk HACIOĞLU
504111312

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Telekomünikasyon Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Bilge GÜNSEL

MAYIS 2014

İTÜ, Fen Bilimleri Enstitüsü'nün 504111312 numaralı Yüksek Lisans Öğrencisi **Canberk HACIOĞLU**, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı **"DERİNLİKLİ ÖĞRENME KULLANILARAK KONUŞMADAN UYKULULUK/UYKUSUZLUK TESPİTİ"** başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı :	**Prof. Dr. Bilge GÜNSEL** İstanbul Teknik Üniversitesi	.....
Jüri Üyeleri :	**Doç. Dr. İlker BAYRAM** İstanbul Teknik Üniversitesi	.....
	Yrd. Doç. Dr. Serap KIRBIZ MEF Üniversitesi	.....

Teslim Tarihi : **5 Mayıs 2014**
Savunma Tarihi : **30 Mayıs 2014**

Babama,

ÖNSÖZ

Lisans ve Yüksek Lisans süresi boyunca beni akademik konularda çalışmaya sevk eden, bana her konuda yardımcı olan değerli tez danışmanım Prof. Dr. Bilge Günsel'e sonsuz teşekkürlerimi sunarım. Çalışmalarında ortaya koydukları öznitelikleri kullanmama izin veren, anlaşılmasında yardımını esirgemeyen Dr. Cenk Sezgin'e ve SLC veritabanını sağlayan Prof. Dr. Jarek Krajewski'ye teşekkür ederim. Benden desteklerini hiçbir zaman esirgemeyen, en az benim kadar bu sürecin stresini çeken anneme; bu tezin oluşmasında bana inanılmaz yardımcı olan kardeşim, yoldaşım Simge'ye ve her konuda olduğu gibi bu konuda da yanımda olan, bana güç katan sevgili Ezgi'ye, akademik çalışmalarıma saygıyla bakan ve destekleyen şirketim Vodafone'a, bu süreçte ihmal ettiğim tüm dostlarıma, akrabalarıma varlıklarından dolayı teşekkür ederim.

Bu tezin oluşabilmesi için gerekli tüm çalışmalar İTÜ Elektronik ve Haberleşme Mühendisliği Bölümü bünyesinde yer alan "Çoğulortam Sinyal İşleme ve Örüntü Tanıma Laboratuvarı"nda gerçekleştirilmiştir. Testlerde kullanılan veri tabanı Wuppertal Universitesi'nden Prof. Dr. Jarek Krajewski tarafından sağlanmıştır.

Mayıs 2014

Canberk Hacıoğlu
Telekomünikasyon ve Bilgisayar Mühendisi

İÇİNDEKİLER

Sayfa

ÖNSÖZ ..... **vii**
İÇİNDEKİLER ..... **ix**
KISALTMALAR ..... **xi**
ÇİZELGE LİSTESİ ..... **xiii**
ŞEKİL LİSTESİ ..... **xv**
ÖZET ..... **xvii**
SUMMARY ..... **xxi**
1. GİRİŞ **1**
1.1 Tezin Amacı 7
1.2 Literatür Araştırması 8
1.3 Hipotez 10
2. ÖZNİTELİK ÇIKARILMASI **11**
2.1 Amaç 11
2.2 Ses Verisine Uygulanan Ön İşlemler 12
2.3 Ses Özniteliklerinin Çıkarılması 16
2.3.1 Normalize Edilmiş Uykululuk Düzeyi Farkı (NUF) 16
2.3.2 Normalize Edilmiş İzgesel Zarf Farkı (NIZD1, NIZD2, NIZD3) 17
2.3.3 Ortalama Uykusuz Blok Sayısı (OUBS) 18
2.3.4 Çerçevenin Seslilik Değeri (ÇSD) 20
2.3.5 Ortalama Harmoniklik Genliği (OHG) 20
2.3.6 Algısal Bant Genişliği 22
2.4 Sözlük Öğrenmesi 23
3. DERİNLİKLİ ÖĞRENME İLE KONUŞMACI DURUM MODELLEME **25**
3.1 Ön-Eğitim Modeli 26
3.2 Kısıtlandırılmış Boltzmann Makineleri ve Karşıtlıklı Iraksay Algoritmaları 27
3.3 Ön-Eğitim ile Ağ Parametrelerinin Öğreticisiz Öğrenilmesi 31
3.3.1 Parametrelerin güncellenmesi 34
3.4 Öğreticili Öğrenme ile Ağ Parametrelerini En İyileme 35
3.4.1 Geri Yayma Algoritmasında Eşlenik Eğim Metodu 37
3.5 Sınıflandırma 38
4. BAŞARIM TESTLERİ **41**
4.1 Uykululuk Konuşma Veri Seti ve Test Senaryoları 41
4.2 Test Başarım Oranları 43
5. SONUÇ VE ÖNERİLER **51**
5.1 Çalışmanın Uygulama Alanı 51
5.2 Tez Sonuçları ve Tartışma 52
KAYNAKLAR **57**
EKLER **59**
ÖZGEÇMİŞ **65**

KISALTMALAR

AFD	**:** Ayrık Zamanlı Fourier Dönüşümü
CD	**:** Karşıtlıklı Iraksay - Contrastive Divergence
ÇSD	**:** Çerçevenin Seslilik Değeri
Hz	**:** Hertz
ITU	**:** International Telecommunication Union
KBM(RBM)	**:** Kısıtlı Boltzmann Makinaları - Restricted Boltzmann Machines
KSS	**:** Kraolinska Uykululuk Ölçeği - Karolinsla Sleepiness Scale
M2M	**:** Makineler Arası İletişim - Machine to Machine Communication
MFCC	**:** Mel-Frequency Cepstrum
NIZD	**:** Normalize Edilmiş İzgesel Zarf Farkı
NSL	**:** Uykusuz - Non-Sleepy
NUF	**:** Normalize Edilmiş Uykululuk Düzeyi Farkı
OHG	**:** Ortalama Harmoniklik Genliği
OUBS	**:** Ortalama Uykusuz Blok Sayısı
SGD	**:** Stokastik Bayırlı İniş - Stochastic Gradient Descent
SL	**:** Uykulu - Sleepy
SLC	**:** Sleepy Language Corpus
STFT	**:** Kısa Zamanlı Fourier Dönüşümü - Short Time Fourier Transform

ÇİZELGE LİSTESİ

Sayfa

Çizelge 1.1 : Literatür ile karşılaştırmada kullanılacak sonuçlar 10
Çizelge 2.1 : Çalışmada kullanılan öznitelikler..... 12
Çizelge 4.1 : Test senaryoları ve senaryolarda kullanılan kümelere dair bilgiler 43
Çizelge 4.2 : Sözlük öğrenme uygulanmış eğitim kümesinde konuşmacı durum sezimi başarım oranları..... 44
Çizelge 4.3 : Sözlük öğrenme uygulanmış genişletilmiş eğitim kümesinde konuşmacı durum sezimi başarım oranları 44
Çizelge 4.4 : İlk test senaryosu, öbeklenmiş eğitim ve öbeklenmemiş test kümeleri başarım oranları 45
Çizelge 4.5 : İlk test senaryosu, eğitimde öbeklenmiş eğitim, testte öbeklenmemiş geliştirme kümesi kullanılarak elde edilmiş başarım oranları 45
Çizelge 4.6 : İlk test senaryosu, eğitimde öbeklenmiş eğitim+geliştirme, testte öbeklenmemiş test kümesi kullanılarak elde edilmiş başarım..... 46
Çizelge 4.7 : İkinci test senaryosu, eğitimde ve testte, aynı kümenin, azaltılmış öbeklenmemiş test kümesi, kullanılması ile elde edilen başarım 46
Çizelge 4.8 : İkinci test senaryosu, eğitimde ve testte, öbeklenmemiş test kümesi, kullanılarak aynı konuşmacının farklı kayıtları testi 47
Çizelge 4.9 : İkinci test senaryosu, eğitimde azaltılmış öbeklenmemiş test, testte öbeklenmemiş geliştirme kümesi kullanılarak elde edilmiş başarım ... 48
Çizgel 4.10: SVM sınıflandırıcısı ile elde edilmiş test sonuçları. 49
Çizelge 5.1 : Literatür ve tez sonuçlarının karşılaştırılması. 54
Çizelge A.1 : Eşlenik eğim algoritmasının özeti 62

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1 : Ön-eğitim ve ayrıştırıcı öğrenme adımlarından sonra iki sınıflı sınıflandırma işlemi. 3
Şekil 1.2 : Çalışmada kullanılan sistemin blok diyagramı. 7
Şekil 2.1 : Öznitelik çıkarımı blok diyagramı, Sezgin, Gunsel, Kurt (2011)'den uyarlanmıştır. 13
Şekil 2.2 : SL(1.sütun)-NSL(2.sütun) sınıflarındaki konuşma kayıtlarından elde edilen ön işlem sonuçları. 16
Şekil 2.3 : NSL ve SL etiketli kayıtlar arasındaki zarf değişimlerinin farkı özniteliğinin kritik banltlarda zamana bağlı değişimi. 18
Şekil 2.4 : NSL ve SL etiketli kayıtlarda uykusuz blokların sayısının kritik banltlarda zamana bağlı değişimi. 19
Şekil 2.5 : NSL ve SL etiketli kayıtlarda seslilik düzeyi farklarının kritik banltlarda zamana bağlı değişimi 20
Şekil 2.6 : NSL ve SL etiketli kayıtlar için uykululuk değişiminin kritik bantlar korelasyonunun zamana bağlı değişimi. 22
Şekil 3.1 : Derinlikli öğrenme blok diyagramı. 26
Şekil 3.2 : Standart KBM ağ yapısı. 27
Şekil 3.3 : Vektörel ve Skaler Gösterilimler. 28
Şekil 3.4 : Gibbs örnekleme işlevi. 34
Şekil 5.1 : Eğitim ve test sırasında alınan hata sayılarının karşılaştırılması. 55

DERİNLİKLİ ÖĞRENME KULLANILARAK KONUŞMADAN UYKULULUK / UYKUSUZLUK TESPİTİ

ÖZET

2000'li yıllardan sonra donanım teknolijisinde yaşanan gelişmeler, yazılım temelli uygulama dünyasının sınırlarını daha önce olmadığı kadar genişletmiştir. Artık herkesin cebinde, elinde taşıdığı yetenekleri azınsanmayacak bir işlem birimi bulunmaktadır. Bu cihazlar insan hayatını kolaylaştırdıkça yayıldı ve dünyadaki beş kişiden biri telefon sahibi oldu. Ekonomik açıdan gelişmiş ülkelerde oran daha da büyümekte olup, Türkiye'de bu oran %30'dur. Makineler arası iletişimin yaygınlaşması ile 2015 yılında 15 milyar cihazın birbirleri ile iletişim halinde olacağı öngörülmekte ve cihaz kümesine artık bilgisayar, telefon, tabletlerin yanında buzdolabı, elektrik sayacı ve arabayı da eklenmektedir. Teknolojide yaşanan bu gelişmeler insan-makine etkileşimi uygulamalarının sayısını, algoritmalarının performans ihtiyaçlarını arttırıcı bir etki yapmakta ve bu uygulamaları günümüzün en çok kullanılan uygulamaları arasına sokmaktadır. Google arama motoru, koca bir metni çevirmemize yarayan tercüme yazılımları ve cep telefonumuza komut göndermenin ötesinde sesli sohbet etme imkanı tanıyan sanal asistan uygulamalarından Siri ilk akla gelen makina öğrenmesi uygulamalarındandır. Fakat araştırmacılar daha ileriyi hedefleyerek, sesten karakter ve duygu analizi yapan uygulamalar, duygusal durumumuza göre hizmet sunan sanal asistanlar ve kendi kendine düşünebilen yapılar kurmak üzere çalışmaktadırlar.

2009 yılında Amerika'da yapılan bir araştırmaya göre yaklaşık 2 milyon kaza uykusuzluk kaynaklı sebeplerden olmaktadır. Bu nedenle son yıllarda ses verisinden otomatik olarak konuşmacının uykululuk / uykusuzluk durumlarını tespit eden ve gerekli uyarı mekanizmalarını devreye sokan sistemler üzerinde çalışılmaktadır. Bu tür sistemlerin tasarımındaki temel zorluklar; uzun süreli modelleme kullanarak hatalı karar olasılığını düşük tutabilme ve gerçek zamanlı karar verebilme şeklinde özetlenebilir.

Bu tez çalışmasında problem olarak ele alınan, konuşmadan uykululuk / uykusuzluk tespiti; insan sesininin probleme uygun özniteliklerinin elde edilerek makinalara öğretilmesi ve öğrenme sonucunda yeni bir örnek geldiğinde konuşmacının uykulu ya da uykusuz durumlarından hangisinde olduğuna karar verilmesi problemidir. Problem, ses işaret işleme ve örüntü tanıma kapsamında iki ana başlıkta incelenerek çözüme gidilmiştir: ses özniteliklerinin elde edilmesi ve konuşmacının durumunun modellenerek sınıflandırılması. Tez çalışmasında, konuşmacının durumunu belirlemek için öncelikle geliştirilen istatistiksel modelin eğitilmesi ve eğitim sonucunda tasarlanan sınıflandırıcının kullanılarak yeni gözlenen ses örneklerinden konuşmacının durumuna karar verilmesi için derinlikli öğrenme kullanılmıştır.

İnsan sesinin modellenmesinde kullanılabilecek öznitelikler iki temel grupta toplanabilir. Birinci grup öznitelikler dilbilimsel (linguistic) olarak adlandırılır ve konuşmacıya ya da konuşmaya bağlı olarak farklılık gösterir. İkinci grup, konuşmayı ses olarak işleyen ve konuşmacının durumunu modellemede kullanılabilecek dolaylı özniteliklerdir. Konuşmacıya ve konuşmaya göre farklılık göstermezler. Bu tez çalışmasında konuşmacı durumlarından uykululuk / uykusuzluk durum sınıflandırması ikinci gruptaki öznitelikler kullanılarak gerçeklenmiştir. Örnek vermek gerekirse, “İyi Geceler !” diyen biri Japon diğeri Türk iki kişinin ne dediği sadece dilleri bilen kişiler tarafından anlaşılabilir, fakat bu cümleyi kurarkenki uykululuk halleri iki dili de bilmeyen kişiler tarafından ayırt edilebilir, hangisinin daha uykulu olduğuna görece bir karar verilebilir. Çalışmada dilbilimsel özniteliklerin kullanılmamasının nedeni, konuşma, konuşmacı tanıma yapmadan konuşmacının durumunu otomatik olarak ve endüşük karmaşıklıkla tespit edebilmektir.

Çalışmada kullanılan öznitelikler literatürde ses analizinde çokça kullanılan bürünsel (prosodic) özniteliklerdir. Ancak çalışmada kullanılan özniteliklerin, yaygın olarak kullanılan bürünsel özniteliklerden farkı, algısal bantta konuşmanın izgesel özelliklerini modelleyebilmesi, aynı zamanda dile ait yapısal özellikleri kullanmadan içeriği izgesel ve zamansal bağlamda takip edebilmesidir. Sesin algısal kalitesini ölçmek amacıyla kullanılan ITU-PEAQ standardı, kullanılan özniteliklerin temelini oluşturmaktadır. Çalışmada kullanılan öznitelik sayısı 9 olup, özniteliklerden 3 tanesi (Ortalama Harmoniklik Genliği, 10 dB Algısal Bant Genişliği, 5 dB Algısal Bant Genişliği) Hertz frekans ölçeğinde, 6 tanesi (Normalize Edilmiş Uykululuk Düzeyi

Farkı, Normalize Edilmiş İzgesel Zarf Farkı(1,2,3), Ortalama Uykusuz Blok Sayısı, Çerçevenin Seslilik Değeri) ise Bark frekans ölçeğinde hesaplanmaktadır.

Eğitim ve sınıflandırmada son dönemde büyük veri modellemede kullanımı önerilen derinlikli öğrenme (deep learning) kullanılmıştır. Bu kapsamda gerçeklenen 3 katmanlı kısıtlandırılmış Boltzmann makinalarının (KBM) ilk katmanı ve ikinci katmanı 500 nörondan oluşur. Son katmanda ise 2000 nöron bulunmaktadir. Kurulan ağ yapısındaki nöronlar arasındaki yönsüz bağlantıların ağırlık katsayıları ve destek terimleri öğrenilerek derinlikli öğrenme gerçeklenmektedir.

Kurulan ağ ile öncelikle ön-eğitim aşaması gerçeklenir. Eğitim ses kayıtları 100'er kayıtlık yığınlar şeklinde ön-eğitim bloğuna verilir. Ön-eğitim başlangıcında ağırlıklara rasgele değerler atanarak ön-eğitim süreci başlar ve KBM'ler üretici (generative) öğrenme ile eğitilir; özyinelemeli olan bu öğrenmenin çalışmada 20 iterasyonda tamamlandığı görülmüştür. Eğitim sonunda elde edilen parametreler eğitimin son adımı olan ayrıştırıcı (discriminative) öğrenmede kullanılır.

Ayrıştırıcı öğrenme aşamasında uykulu / uykusuz olarak etiketlenmiş eğitim kayıtları kullanılır. Ön-eğitim aşamasında elde edilen değerler geri-yayma (backpropagation) algoritması kullanılarak özyinelemeli olarak iyileştirilir. Derinlikli öğrenme çok parametreli bir eniyileme problemi olarak modellenir ve ağ parametrelerinin hiyerarşik bir yapıda öğrenilmesi amaçlanır. Eniyilemede kullanılan geri yayma prosedürü temel olarak istenilen çıkış vektörü ile ağın ürettiği çıkış vektörü arasındaki karesel hatayı özyinelemeli olarak en küçüklemeyi amaçlar. Ağırlıklara getirilen kısıtlar altında gerçeklenen içbükey eniyileme ile girişte ya da çıkışta yer almayan ağın hiyerarşik yapısı içerisindeki saklı işlem birimleri, konuşma durumuna ilişkin önemli öznitelikleri yansıtmaya başlarlar. Prosedür en alt katmandan başlayarak katman içi birimlerin durumlarını paralel olarak, üst katmanlara doğru ise seri olarak günceller ve eniyilemeyi gerçekler. Hata fonksiyonunun en küçüklemesi işleminde, bu çalışmada hızlı ve basit olması nedeniyle eşlenik eğim (*conjugate gradient*) algoritması en iyileme metodu olarak kullanılmıştır. 20 iterasyonun her birinde ayrıştırıcı öğrenme ile en iyilenen ağın parametreleri kullanılarak test kümesi içerisinde yer alan vektörler sınıflandırılır ve her iterasyonda sınıflandırma hataları ortaya konur. Eğitim ve sınıflandırma hatalarının birlikte azalması, kurulan ağın eğitim-sınıflandırma performansını gösteren bir metrik olarak izlenir.

Tez çalışmasında literatürde kullanılan sonuçlarla karşılaştırma yapmak amacıyla SLC (Sleepy Language Corpus) veri tabanı kullanılmıştır. Sınıflandırma sonuçları literatürde yapılan benzer çalışmaların sonuçları ile karşılaştırılmıştır. Derinlikli öğrenme ile eğitim kümesi içerik ve büyüklük olarak değiştikçe uykululuk "SL" ve uykusuzluk "NSL", ikili sınıflandırma sonucundaki performansın korunduğu gösterilmiştir. Referans alınan çalışmalardaki iki senaryo üzerinden sonuçlar raporlanmış olup ilk senaryo için literatürde ağırlıklandırılmamış ortalama hatırlama oranları üzerinden raporlanan taban seviye olan %67.3'e karşılık %56'lık, ikinci senaryo için ise %70.3'e karşılık %74'lük başarım oranı elde edilmiştir.

SLEEPINESS DETECTION FROM VOICE BY USING DEEP LEARNING SUMMARY

Developments in technology especially in hardware technology after early 2000s, enhanced the borders of software application world. People of mobilized world has highly capable processing units in their pockets and in their bags. More people than yesterday have started to use the smart devices because of their contribution to daily life and business. One of five people in the world has a smart device and the ratio is even more than that in economically developed countries, e.g. for Turkey the penetration is around %30. On the orher hand, numbers don't represent the peak values of trend; after machine to machine business gets maturity level in the market, the penetration will rise and there will be 15 billion connected devices according to forecasts. Another major improvement is in the set of connected devices, after Machine-to-Machine (M2M) communications fridges, televisions, electric meters, combis, cars and other helpful appliances in human daily life are adding to set and becoming ready to interact with human society. These developments in technology make applications using that devices more important and most commonly used; search engine Google, online translation tools capable of translating whole document and virtual assistants like Siri that make us to chit-chat with them is prominent machine learning applications. As proposed, these are not the limits of machine learning world, popular topics in machine learning focusing on human state detection i.e. affective computing like character analysis or emotion detection from speech, self-learning structures.

According to results of survey held in USA, the cause of two million car accidents is driver sleepiness. Researchers are taking these kind of problems as a subject of affective computing and putting forth research outputs to solve them. If the cars were able to detect driver sleepiness level, drivers reached the sleepiness threshold wouldn't start their car engine and probably car accidents wouldn't be faced.

In this thesis, the aim is to detect sleeepines with nine perceptually masked prosodic features trained and classified by Deep Learning algorithms.

The problem of sleepiness detection is evaluated in two parts. In the first part proper features of human voice are extracted to teach machines. In the second part proper features are leaned to model the structure with the aim of classification of records. The classification results determine the state of the speaker in terms of sleepiness.

Human voice carries two different contents over two different channels. The first channel is the direct channel and the content reflects "what we are saying". The second channel is the indirect channel, channel carries the honest signals and the content reflects "how we are saying". The content of the first channel is very highly and strictly dependent on to the linguistic properties of spoken language. However the honest signals carried over the second channel is independent from the speaker or the spoken language, and gives the state of the speaker. For example, when a Japanese and a Turkish say "Good Evening !" before sleeping, only people who speak their language understand "what they are saying". On the other hand, all people, including the people who can't speak their languages, can evaluate their sleepiness level with the help of "how they are saying". Therefore selected features should represent second type of the content very well to decide the correct state.

The speech features are extracted from *"SLC Corpus"* published for *"Interspeech 2011 Speaker State Challenge"*. The features used in thesis are based on the prosodic features which are also used in the literature in the field of similar subjects, like speech recognition and speaker state detection. Conventional features are re-formulated to represent sleepiness levels, prosodic features measuring the loudness differences at different levels rather than loudness itself as in the conventional methods. Furthermore features used are capable of monitoring the loudness differences not only through the critical bands but also in the temporal domain. The essentials of the features are coming from the International Telecommunication Union's standard ITU-PEAQ. Nine features are used in the research to model the speaker state content, three (Average Harmonic Structure Magnitude, 10dB and 5 dB Perceptual Bandwidth) of the used features are in Hertz frequencies and six (Normalized Sleepiness Level Difference, Normalized Spectral Envelope Difference 1-2-3, Average Number of Non-Sleepy Blocks, Overall Loudness of the frame) of them in Bark scale.

Applied training and classification methods are based on Geoffrey Hinton's proposed algorithms under the concept of Deep Learning. The learning consists of two phases;

first phase is pre-training and the second phase is fine-tuning. Pre-trainining which is the generative learning of the model is initialized by random parameters (weights between hidden and visible units, and bias term) and generates the weights and the biases for discriminative learning phase which is fine-tuning. Restricted Boltzmann Machines (RBM) are used for the pre-training, with three layers. The first and second layers consist of 500 hidden units and the third layer has 2000 hidden units. Starting from first batch of feature vectors, deep RBM network successively and iteratively learns the data and represents it with parameters. Since the algorithm is based on matrix multiplications, under the constraint of computational complexity, all calculation are based on batches with 100 records.

After pre-training have generated the initial parameters for discriminative learning, the last stage of training which is fine-tuning starts. In contrast to the generative learning, the discriminative learning needs labelled data. Since we know the desired output, the parameters estimated in the pre-training stage are fine-tuned to give the desired output. The fine-tuning algorithm used in thesis is the backpropagation algorithm. The backpropagation algorihm aims to minimize the difference between the actual and desired output. Starting from the first layer backpropagation algorithm optimizes the weight bottom-to-top and iteratively (same iteration number used with training of RBM). Conventionally, optimization of that kind of error function is based on the gradient descent methods. Similarly, *"Conjugate Gradient Descent"* algorithm is used in this work as minimization algorithm.

Classification of the test set starts with backpropagation classification error is reduced iteratively using the optimized weights and biases. Classification results reported over using similar test scenarios in the literature. The success is to keep consistency between different test sets varies in terms of length and content, and to keep consistency between two class, "SL" and "NSL". Recent works used as reference, report performance over two different scenarios. By using same training and test sets in these studies, comparable results generated as an output of the thesis. Baselined success ratio for recall rates for the first scenario and the second scenarios are 67.3% is 70.3%, respectively results gathered from thesis for the first scenario and the second scenario is is 56% and 74%. Beyond the comparable results, result section takes into account some other metrics for the success of the proposed work and satisfied results are gathered from different sets.

1. GİRİŞ

İnsanların etkileşimlerini bugüne kadar yapılmış çalışmalar ışığında iki ana iletişim kanalı üzerinden yaptıklarını söyleyebiliriz [1]. Doğrudan kullanılabilir bilgilerin ve dolaylı bilgilerin iletildiği bu iki farklı iletişim kanalının özellikleri kullanılarak geliştirilen insan-makine etkileşimi uygulamaları, farklı amaçlar için insanlığın hizmetine sunulmaktadır. İnsanların çevreyle ilişkilerinde bu kanalların kullanımlarını örnekleyebiliriz; bir arkadaşımızı gördüğümüzde "Nasılsın ?" sorusuna aldığımız yanıt genelde "İyiyim !" olur. Kişinin burada **"ne"** söylediği açıktır fakat **"nasıl"** söylediği cevabı alan kaynağı ya tatmin eder ya da etmez ve cevabın tatmin etmediği durumda kaynak "**kişinin gerçek durumunu"** anlamak için sorularına devam eder. Çalışmaya da uygun olan başka bir örnekte bir Japon ve bir Türk turistin kendi dillerinde **"iyi geceler!"** dediği durumu ele alalım. Direk kanalı kullanarak geliştirdiğimiz uygulamalar ile, simultane çevirisini yaparak kişinin **"ne"** dediğini günümüz teknolojisi ile kolayca anlayabiliyoruz; fakat **"nasıl"** ifade ettiği bilgisini taşıyan dolaylı kanal kullanarak uykululuk seviyesinin ne olduğunu bu kadar kolay anlayamıyoruz. Örneklerden hareketle birinci kanal ile ikinci kanal arasındaki fark basitçe ortaya konulabilir; direk kanal kişinin **"ne söylediği"** ile ilgilenirken, dolaylı kanal söylenen şeyin **"nasıl söylendiği"** ile ilgilenir.

Ses ile iletişimde ikinci kanal yani dolaylı bilgilerin taşındığı kanal kullanılarak çeşitli insani durumlar problem olarak ele alınabilir. Bu tezde ele alınan uykululuk/uykusuzluk probleminin [2] yanında, alkollülük tespiti [3], duygu kestirimi [4], liderlik tespiti [5] gibi problemler literatürde hala güncelliğini koruyan popüler örneklerdir. Literatürde duygular hakkında pragmatik bir karar verilip en iyi deneyim olarak referans gösterilen çalışmaların gerçekte uygulamaya dönüşememesinin birkaç sebebi vardır [1]. İlk problem, dünyanın farklı bölgelerinden farklı grupların gerçekleştirdiği çalışmaların ortak bir deney kümesi üzerinde çalışıp ortak sonuçlar ortaya koyamamasıdır. Bu problemi aşmak için, araştırmacıların aynı veri tabanları üzerinde en iyi sonucu almak için yarıştığı akademik yarışmalar [6] önem arz etmektedir. Bu çalışmada ele alınan deney

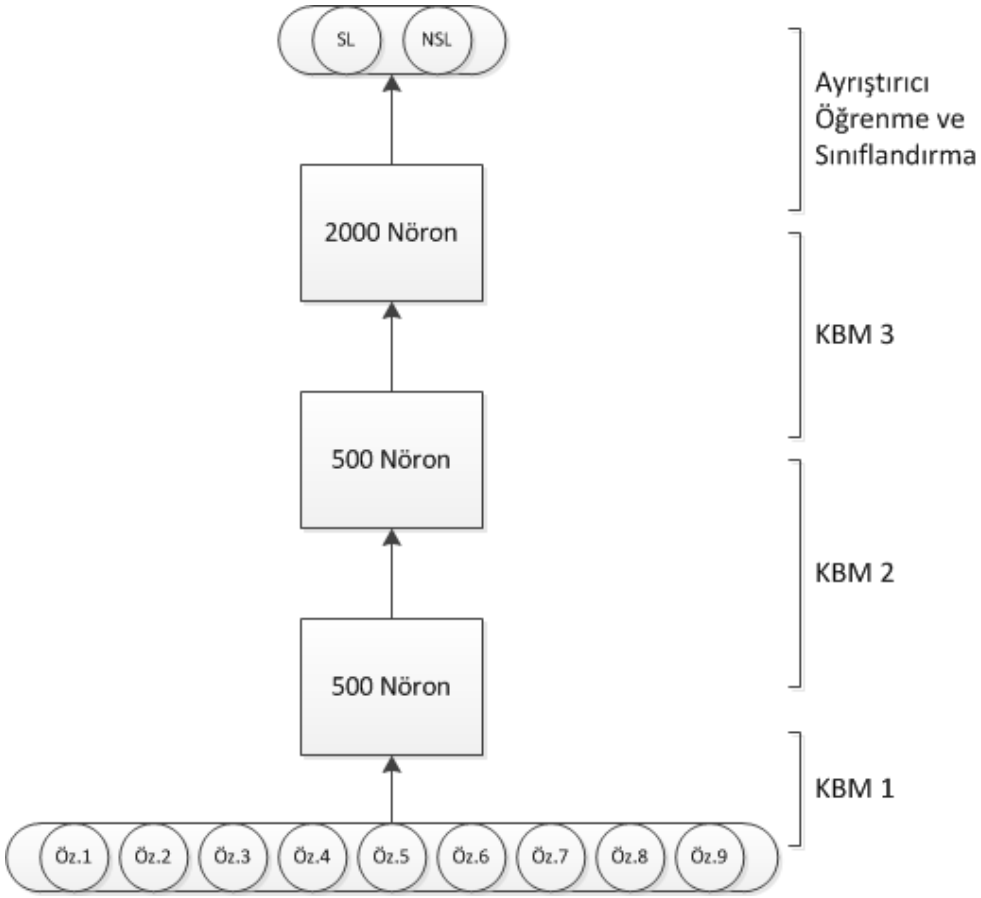
kümeleri, 2011 yılında düzenlenen bir akademik yarışma olan "Speaker State Challenge" [6] [7] için sağlanan ve yarışma kapsamında birçok araştırmacının sonuç gösterdiği deney kümeleridir. İkinci problem, insani durumların kültürler arasında gösterdiği farklılıklardır. Özellikle duygu analizi problemlerinde yaşanan bu sorunun aşılabilmesi için seçilen özniteliklerin konuşmacıdan ve konuşmadan bağımsız varlık gösterebilen öznitelikler olması gerekmektedir. Bu çalışmada kullanılan öznitelikler [4] ispatlanmaya çalışıldığı üzere bu bağımsızlığı sağlamada başarı göstermektedir. Üçüncü olarak insanlar duygularını bastırmak ve amaca uygun olarak değiştirmek için çaba sarf ederler ve başardıkları durumlar da yaşanabilir. Çözümün arabalarda kullanıldığı durumda, yakın gelecekte piyasaya çıkması muhtemel [8] bir çözüde kişinin ideal şartları sağlayamadığı durumda arabasını çalıştıramaması konuya iyi bir örnek olacaktır. Öznitelikler doğru seçilmediği ve örnek sayısı arttırılarak gerçek zamanlı sınıflandırma yapılamadığı durumda, kurulan sınıflandırıcı sistemin hata yapması muhtemel olacaktır. Bu tezde önerilen katmanlı eğitim ve sınıflandırıcı yapısı [9] insan beyninin çalışmasına benzerlik göstermesi nedeniyle problemin karmaşık yapısına daha uygun olacağı kabulüyle seçilmiştir.

Bölüm 4'te detaylandırılacak olan bitirme tezi çalışmasında kullanılan kayıtlar [6] [7], üç ana kayıt kümesinden oluşur: eğitim, geliştirme ve test veri kümeleri. Eğitim veri kümesinde 20 kadın, 16 erkek olmak üzere toplam 36 konuşmacının oluşturduğu 3366 kayıttan elde edilen ses çerçeveleri, test veri kümesinde 19 kadın 14 erkek olmak üzere toplam 33 konuşmacının oluşturduğu 2808 kayıttan elde edilen ses çerçeveleri, geliştirme veri kümesinde ise 17 kadın 13 erkek olmak üzere toplam 30 konuşmacının oluşturduğu 2915 kayıttan elde edilen ses çerçeveleri bulunur. Senaryolara göre gerçekleştirilecek deneyler bu 3 farklı kümenin ve öbeklenmiş hallerinin farklı kombinasyonları ve birleşimleriyle gerçekleştirilir. Kayıtların hepsi gerçek problemlere uygun olması açısından sınıf ortamında ya da araba içerisinde gerçekleştirilmiştir ve kayıtlar KSS (Karolinska Sleppiness Scale) ölçeğine [10] göre 10 farklı uykululuk seviyesi içermektedir.

Kurulan eğitim-sınıflandırıcı yapısının ilk bloğu öznitelik çıkarım bloğudur. Benzer problemlerde de karşılaşılabileceği üzere, çalışmada kullanılan öznitelikler [4] bürünsel özniteliklerdir. Fakat çalışmada kullanılan özniteliklerin, yaygın olarak kullanılan bürünsel özniteliklere göre üstünlüğü algısal bantta konuşmanın izgesel özelliklerini modelleyebilmesi, aynı zamanda dile ait yapısal özelliklerle

uğraşmadan içeriği zamansal bağlamda takip edebilmesidir [4]. Sesin algısal kalitesini ölçmek amacıyla kullanılan ITU [11] standardı, kullanılan özniteliklerin temelini oluşturmaktadır. Çalışmada kullanılan öznitelik sayısı dokuz olup, Bölüm 2'de detaylandırılacak özniteliklerden üç tanesi (Ortalama Harmoniklik Genliği, 10 dB Algısal Bant Genişliği, 5 dB Algısal Bant Genişliği) Hertz frekans ölçeğinde, kalan altı tanesi (Normalize Edilmiş Uykululuk Düzeyi Farkı, Normalize Edilmiş İzgesel Zarf Farkı(1,2,3), Ortalama Uykusuz Blok Sayısı, Çerçevenin Seslilik Değeri) ise Bark frekans ölçeğinde hesaplanmaktadır.

Şekil 1.1'de görülen ve Bölüm 3'te detaylandırılacak olan eğitim-sınıflandırıcı yapısının 2. ve 3. blokları, ön-eğitim ve ayrıştırıcı öğrenme adımlarıdır. Eğitim ve sınıflandırmada kurulan yapı üç katmanlı bir derinlikli sinir ağıdır [9]. İlk katman ve ikinci katmanda 500 nöron bulunan ağın son katmanında 2000 nöron bulunmaktadır. Ön-eğitim, sınıflandırma işleminde kullanılacak ağ parametrelerinin rasgele seçilip kullanılmasını engellemek için yapılan, parametrelerin en iyi sınıflandırıcı performansı için iyileştirilmesine dayanan bir ön işlemdir.



Şekil 1.1 : Ön-eğitim ve ayrıştırıcı öğrenme adımlarından sonra iki sınıflı sınıflandırma işlemi.

Ön-eğitim işleminde kullanılan Kısıtlandırılmış Boltzmann Makinaları (KBM) [12] ile ön-eğitim aşamasında ve derinlikli öğrenmenin diğer aşamalarında her bir konuşma için oluşturulan dokuz boyutlu öznitelik vektörleri tek tek işleme alınmaz, matris işlemlerine dayanan derinlikli öğrenmede 100 vektörün(bu çalışmada ses çerçeveleri) herhangi bir sıra gözetmeden rastgele bir araya getirilmesiyle yığınlar oluşturulur ve bu yığınlar üzerinden işlemler gerçekleştirilir.

Kısıtlandırılmış Boltzmann makinaları, etiketlenmiş veriye ihtiyaç duymadıkları için ön-eğitimde tercih edilen en yaygın metotlardan biridir. İlk KBM'nin rasgele seçilmiş parametreler ile eğitilmesinden sonra bir önceki KBM'nin çıkışı bir sonraki KBM'nin girişi olacak şekilde üç KBM'nin ard arda eğitilmesi sonucunda KBM eğitimi tamamlanır. KBM eğitim işlemi tanımlanan bir üstel enerji fonksiyonunun enküçükleme problemi olarak ele alınır ve seçilen hata fonksiyonu olan "negatif logaritma olasılık fonksiyonunun" en küçüklenmesi ile gerçekleştirilir. Hata fonksiyonunu en iyilemek etmek için *"Stochastic Gradient Descent (SGD)"* algoritması kullanılır ve en iyileme için kullanılacak fonksiyon iki kısımdan oluşur: pozitif ve negatif faz. Pozitif faz gözleme bağlı değişirken, negatif faz sadece modele bağlıdır. Pozitif fazda saklı işlem birimleri üzerinden "beklenen kısmi türev" değeri, gözlemler koşulu altında enerjinin parametreler ile kısmi türevinin alınması ile bulunur. Negatif fazda ise, görünür işlem birimleri ve saklı işlem birimleri üzerinden modeldeki enerji fonksiyonunun kısmi türevinin beklenen değeri hesaplanır. Beklenen ifadesinin kullanılmasının nedeni saklı değerlerin bilinmemesidir. Negatif fazdaki bu beklenen değerin hesaplanabilmesi için her KBM'nin eğitilmesinde Geoffrey Hinton tarafından geliştirilen *"Contrastive Divergence"* [13] algoritması kullanılır ve sonuçların iyileştirilmesi amacıyla 20 iterasyon yapılır. İterasyon sayısı değiştirilebilir bir sayı olup, çalışmada elde edilen sonuçları değiştirmeyecek ve işlemsel karmaşıklık-performans dengesini sağlayacak şekilde 20 seçilmiştir. 3 KBM'nin art arda iterasyonlarla eğitilmesi sonucunda "üretici öğrenme" işlemi sonlanır ve ayrıştırıcı öğrenme sırasında kullanılacak destek terimleri ile ağırlık katsayı matrisleri elde edilmiş olur.

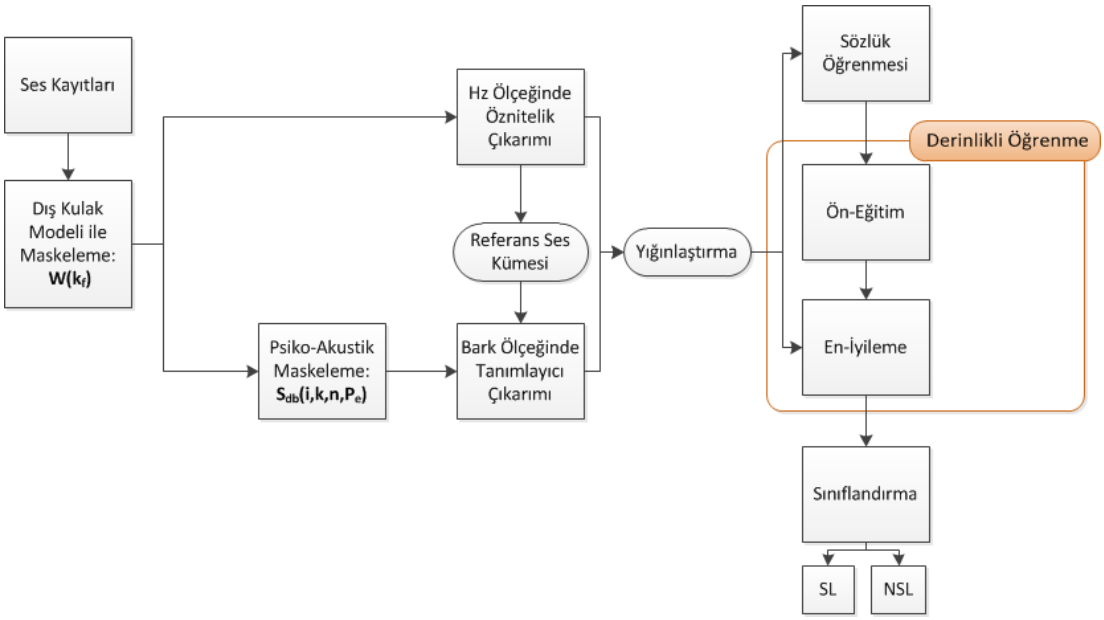
KBM'ler aracılığıyla parametrelerin öğrenilmesi ile sadece üretici öğrenme kısmı tamamlanmaktadır. Ağın eğitim kısmı ayrıştırıcı öğrenme ile tamamlanacaktır. Ayrıştırıcı öğrenme etiketlenmiş veriye ihtiyaç duyan ve ön-eğitim aşamasında elde edilen parametreleri iyileştirerek sonsal olasılıklara çeviren eğitim süreci için bir

iyileştirme ve son basamakta test kümesi için bir sınıflandırma metodu olarak kullanılmıştır [9]. Ağın eğitim aşamasının son basamağında kullanılan geri-yayma algoritması aracılığıyla ağın istenilen sonuçları üretebileceği parametrelere yakınsamasını sağlar. Geri-yayma algoritması, ilk adımda üretici bir model ile elde ettiğimiz başlangıç parametrelerini, ayırıcı modelin başlangıç değerleri olarak kullanmakta ve gözetimli eğitim gerçekleştirmektedir. Eniyilemede kullanılan geri yayma prosedürü temel olarak istenilen çıkış vektörü ile ağın ürettiği çıkış vektörü arasındaki farkı en küçüklemeyi amaçlar. Bu işlem tekrarlı olarak gerçekleşir. Ağırlıklara getirilen düzenlemeler ile girişte ya da çıkışta yer almayan ağın içerisindeki saklı işlem birimleri önemli öznitelikleri yansıtmaya başlarlar. Prosedür en alt katmandan başlayarak katman içi birimlerin durumlarını paralel olarak, üst katmanlara doğru ise seri olarak günceller ve eniyilemeyi gerçekleştirmeyi amaçlar [14]. Hata fonksiyonunun en küçükleme işleminde, bu çalışmada hızlı ve basit olması nedeniyle eşlenik eğim (*conjugate gradient*) algoritması en iyileme metodu olarak kullanılmıştır [15]. En küçükleme işleminde kullanılan kod ve algoritma Carl Rasmussen tarafından geliştirilen en küçükleme algoritmasıdır [16]. 20 iterasyonun her birinde ayrıştırıcı öğrenme ile en iyilenen ağın parametreleri kullanılarak test kümesi içerisinde yer alan vektörler sınıflandırılır ve her döngüde sınıflandırma hataları ortaya konur. Eğitimde yaşanan hatalar ile sınıflandırmada yaşanan hataların birlikte azalması kurulan ağın eğitim-sınıflandırma performansını gösteren bir metrik olarak ele alınacaktır.

Sonuçlar dört ana kriter üzerinden değerlendirilecektir. Kurulan ağın doğrulanması için yapılan ilk testler kayıtların hem test hem de eğitimde kullanıldığı durum için gerçekleştirilmiştir ve literatürdeki adı ile "*train vs train*" performansı raporlanmıştır. Testler sonucunda hatırlama oranı için yaklaşık %67'lik ağırlıksız ortalama sonucu elde edilmiştir. Doğrulama testlerinden sonra iki ana test senaryosu izlenmiş olup her bir senaryo için farklı amaçlar ile testler gerçekleştirilmiştir. İlk senaryo ile ikinci senaryo arasındaki fark ilk senaryoda eğitimde öbekleme işlemine tabi tutulmuş küme kullanılırken; ikinci senaryoda eğitimde öbeklenmemiş kümelerin kullanılmasıdır. Burada amaç referans çalışmalarda [17] öbeklenmiş küme ile alınan daha yüksek sonuçlarla, tez çalışmasında aynı senaryolar üzerinden elde edilen sonuçların karşılaştırılmasıdır. İlk senaryo amacı dahilinde kurulan ağın performansını göstermede kullanılacak ilk ana kriter eğitim ve testte farklı kayıtlar ve

kişiler kullanılmasıyla başarım oranlarının göstereceği tutarlılık olacaktır. İkinci ana kriter *"SL"* ve *"NSL"* olarak raporlanacak olan *"uykululuk"* ve *"uykusuzluk"* performansları arasında tutarlılık olacaktır. Üçüncü kriter ise eğitim kümelerinin büyümesi ile sonuçların arasındaki tutarlılığın korunması olacaktır. Üç kriter de ilk senaryo üzerinden incelenirse; ilk kriterin incelenmesi için eğitimde ve testte (testte kullanılan küme literatürde en zor küme olarak kabul edilen kümedir) farklı kayıt ve konuşmacıların kullanıldığı test kurgusu sonucunda ağırlıklandırılmamış hatırlama oranı %56 çıkmış olup, "SL", %61, ve "NSL", %52, hatırlama oranları birbirine oldukça yakındır ki bu da ikinci kriterin sağlandığını göstermektedir. Daha kolay bir test kümesi ile sonuçlar tekrarlandığında ağırlıklandırılmamış hatırlama oranı %70'e ("SL", %58, ve "NSL", %83 ağırlıklandırılmamış hatırlama oranları ile) çıkmaktadır. Üçüncü kriterin sağlanması için eğitim kümesi büyütülerek sonuçlar tekrarlanırsa, önceki kurguda %70 çıkan ağırlıklandırılmamış hatırlama oranının %74 çıktığını, ve "SL" hatırlama oranının değişmeyerek %58 kaldığını, "NSL" oranının ise yükselerek %90'a çıktığını görüyoruz. Bu üç kriter ışığında test senaryosu-1'in kurulan sistemin gerçek probleme uygunluğunu ispatladığı sonucuna varılır. İkinci test senaryosu ile performansın arttırılması konusu ele alınacak olup, dördüncü kriter olan öbeklenmemiş kümeler ile sonuçlarda yaşanan değişiklikler metrik olarak kullanılacaktır. Eğitim ve testte farklı kümelerin kullanıldığı ilk kurgu için ağırlıklandırılmamış ortalama hatırlama oranı %64 olup "SL" ve "NSL" sınıfları için %70 ve %58 ağırlıklandırılmamış hatırlama oranı elde edilir. Aynı konuşmacıların farklı kayıtlarının yer aldığı, aynı veriyi içermeyen, eğitim ve test kümeleri kullanıldığında ağırlıklandırılmamış ortalama hatırlama oranı %82'ye ("SL" %89 ve "NSL" %75 ağırlıklandırılmamış hatırlama oranları ile) çıkmaktadır.

Test senaryoları ve senaryolar içerisindeki kurgular ile modellenen sistemin probleme uygunluğu gösterilmiş olup, en zor kümeler kullanılarak minimum performans çıktıları ortaya konmak istenmiştir. Bunun yanında kümelerin değiştirilmesiyle yaşanan performans iyileşmeleri probleme uygun daha yüksek performanslı modellerin kurulabileceğini göstermektedir.



Şekil 1.2 : Çalışmada kullanılan sistemin blok diyagramı.

1.1 Tezin Amacı

Makina öğrenmesi algoritmaları kullanılarak oluşturulan uygulamalar 2000'li yıllarda maliyet-performans kısıtı altında insan hayatını en çok kolaylaştıran uygulamalar olarak sayılabilir ve şüphesiz ki ilk akla gelen 1990'ların sonu 2000'lerin başında doğan Google arama motoru olacaktır. Devam eden yıllarda koca bir belgeyi tek tuşla istediğiniz dile çevirmeye yarayan tercüme uygulamaları, telefonunuza sadece komut göndermekle kalmayıp sohbet edebilmenizi sağlayan sanal sekreter uygulamaları örnekleri çoğalarak performans açısından tatmin edici bir noktaya gelmektedir. Performans beklentisinin ve performanslı uygulamaları geliştirmeye, kullanmaya olanak sağlayan donanım-yazılım olanaklarının artması araştırmacıları daha iyi modeller bulmaya sevk etmiştir.

Bu alandaki daha iyi sistem tasarım çalışmaları üç farklı alana yoğunlaşmıştır. İlk alan probleme konu olan verinin en iyi şekilde verideki bilgiyi kaybetmeden dijital yansımasının elde edilebilmesi, yani daha iyi özniteliklerin keşfedilmesini konu alır. İkinci çalışma alanı veriyi ya da dijital yansımasını en iyi şekilde öğrenip sınıflandırılmasını sağlayacak öğrenme ve sınıflandırma algoritmalarının geliştirilmesini konu alır. Üçüncü çalışma alanı ise bu uygulamaların ve algoritmaların geliştirilebilmesini engelleyen donanım ve yazılım mimarisindeki teknolojik kısıtları ortadan kaldırmaya odaklanır.

Giriş bölümünde belirtildiği gibi tasarlanan sistemin öznitelik çıkarım basamağı daha önce uykululuk tespit probleminde kullanılmıştır [17]. Aynı problemde olmasa da eğitim ve sınıflandırma algoritmaları insani durum tespitinde [18] kullanılmıştır. Tez çalışması ile amaç, bir insan-makine etkileşimi problemi olması dolayısıyla makina öğrenmesi alanına giren konuşmacı durum tespiti, daha özelde "sesten uykululuk tespiti" probleminin çözümü için daha önce kullanılmamış bir sistem tasarımı sunarak, eğitim ve sınıflandırmada derinlikli öğrenme kullanılarak, literatüre elde edilen sonuçlar ile katkı yapmaktır.

1.2 Literatür Araştırması

İnsan sesini kullanarak geliştirilen makine öğrenmesi algoritmaları kullandıkları öznitelikler bakımından ikiye ayrılırlar. İnsan sesi iki farklı kanal tarafından taşınan dilbilimsel ve duygusal özellikler olmak üzere iki farklı enformasyon içerir [1]. Birinci kanal ile dilbilimsel özellikleri kullanarak geliştirilmiş uygulamalar [19], ikinci kanal kullanılarak geliştirilen uygulamalara göre daha fazladır. Genelde konuşma ve konuşmacı tanıma temelli problemleri konu alan bu çalışmalarda *"Saklı Markov Modeller"* sıklıkla kullanılmış olup başarılı çalışmalara imza atılmıştır [20].

Bilgisayar bilimleri ile sosyal bilimlerin buluştuğu hesaplamalı sosyal bilimler alanındaki önemli çalışmalarda kullanılan ikinci kanal henüz gelişmekte olan bir çalışma alanı olarak gösterilmektedir ve insani durum tespiti problemlerinde kullanılmaktadır. Sadece insanlara ait olmayan bu kanal üzerinden canlılığın devamı için hayati öne sahip birçok sinyalin taşındığı bilinmektedir [21]. Kuşların tehlike anında çıkardıkları, kendi türleri ve avcılar tarafından sinyal olarak algılanan sesler, insanların sosyal ortamlarındayken çevreye yaydıkları sesler içerisindeki bazı bileşenler dürüst sinyaller olarak adlandırılır. Dürüst olarak adlandırılmalarının nedeni ise doğanın işleyişine uygun olarak her zaman dürüst bir şekilde durumları ortaya çıkarmalarıdır. Literatürde de amaç yüksek performans için sayısal verilerden dürüst sinyalleri temsil eden en iyi öznitelikleri bulmaktır.

İnsani durum tespiti, insan-makine etkileşimi konularından kolaylıkla çözülmesi mümkün olmayan problemlere [1] sahip zor sınıftadır. İlk problem dünyanın farklı bölgelerinden farklı grupların gerçekleştirdiği çalışmaların ortak bir deney kümesi üzerinde çalışıp ortak sonuçlar ortaya koyamamasıdır. İkinci problem insani durumların kültürler arasında gösterdiği farklılıklardır ki özellikle duygu analiz

problemlerinde yaşanan bu sorunun aşılabilmesi için seçilen özniteliklerin konuşmacıdan ve konuşmadan bağımsız varlık gösterebilen öznitelikler olması gerekmektedir. Üçüncü olarak insanlar duygularını bastırmak ve amaca uygun olarak değiştirmek için çaba sarf ederler ve başardıkları durumlar da yaşanabilir. Çözümün arabalarda kullanıldığı durumda, yakın gelecekte piyasaya çıkması muhtemel [8] bir çözüde kişinin ideal şartları sağlayamadığı durumda arabasını çalıştıramaması konuya iyi bir örnek olacaktır. Öznitelikler doğru seçilmediği ve örnek sayısı arttırılarak gerçek zamanlı sınıflandırma yapılamadığı durumda kurulan sınıflandırıcı sistemin hata yapması muhtemel olacaktır. İnsani durum kestirimi çalışmalarına kaynak oluşturan farklı veri kaynakları vardır ve en popülerleri sesin [2] ve yüzün [1] [6] farklı özelliklerini kullanan kaynaklardır.

Çalışmada ele alınan "uykululuk tespiti" probleminin [2] yanında sesten farklı insani durumlar tespit edilebilir. Bunlardan "alkollülük tespiti" [3], "duygu kestirimi" [4], "liderlik tespiti" [5] gibi problemler literatürde hala güncelliğini koruyan popüler örneklerdir. Sesi kaynak olarak alan araştırmaların çoğunda bürünsel öznitelikler ve istatistiksel özellikleri öznitelikler [6] olarak kullanılır. Bunlardan temel ve perde frekanslar, enerji, seslilik oranı, Mel-frekans spektral bileşenleri en bilinenleridir. Münih Teknik Üniversitesi tarafından geliştirilen "OpenEar" yazılımı [22] kullanılarak 4000'den fazla öznitelik elde edilebilir.

Öznitelik çıkarımından sonra bu öznitelikleri kullanacak eğitim ve sınıflandırma algoritmaları başarımı belirleyen önemli unsurlardan diğeridir. Uykululuk tespiti problemlerinde kullanılan "Destek Yöney Makinaları" [17], "Yapay Sinir Ağları" [2] ve bunlardan oluşan hibrid sınıflandırıcılar yaygın olarak kullanılan algoritmalardır. Uykululuk tespitinde daha önce kullanılmamış olsalar da insani durum tespiti araştırmalarında derinlikli öğrenme algoritmaları daha önce kullanılmıştır [19] [18].

Derinlikli öğrenme algoritmaları farklı tasarımlarla sınıflandırma ve öğrenim algoritmalarında kullanılmışlardır ve birçok derinlikli öğrenme algoritması Geoffrey Hinton ve Toronto Üniversitesi'ndeki grubunun gerçekleştirdiği çalışmalara dayanmaktadır. Derinlikli öğrenme bir konsept olarak farklı algoritmaların kullanıldığı farklı yapıları kapsayan çatı bir konsepttir. Temel olarak katmanlı sinir ağları ve katmanlı graf tabanlı modeller (Derinlikli İnanç Ağları, derinlikli Boltzmann Makinaları) olarak ikiye ayrılabilir. Tez çalışmasında KBM'ler ön-eğitim basamağında kullanılmışlardır fakat farklı çalışmalarda KBM'lerin

konumlandırılması farklı olmuştur. Sadece eğitim aşamasında değil hem eğitim hem sınıflandırma aşamasında ayırıcı model olarak kullanılan derinlikli KBM ağları [23] [24], bu çalışmada da kullanıldığı şekilde ön eğitim aşamasında üretici model olarak kullanılan KBM ağları [9] ve hem üretici hem ayırıcı modelin kullanıldığı hibrit yapıları [19] kullanan çalışmalar örnek olarak verilebilir.

Derinlikli öğrenme yapılarının farklılaştığı bir diğer nokta, geri yayma algoritmalarıdır. Geri yayma algoritmalarında geleneksel olarak en küçükleme için *"Bayır İnişi (Gradient Descent)"* algoritmalarını kullanır [14], fakat bu çalışmada da olduğu gibi son dönemde daha etkin en küçükleme algoritmaları olduğu ispatlanmıştır [15] [25].

Geri yayma algoritması etiketlenmiş veriye ihtiyaç duyduğu için üretici modeller ile sınıflandırma yapan bazı çalışmalarda kullanılmayarak yerine Hinton tarafından geliştirilen bir diğer algoritma olan *"Wake and Sleep"* [26] algoritması kullanılmıştır.

Uykusuzluk tespiti probleminin literatürde referans alınan sonuçları Çizelge 1.1'de belirtilmiştir. Bitirme çalışmasında aynı kümeler kullanılarak karşılaştırmak üzere sonuçlar elde edilmiş ve bu sonuçlar ile son bölümde karşılaştırmaya gidilmiştir.

Çizelge 1.1 : Literatür ile karşılaştırmada kullanılacak sonuçlar.

Çalışmalar	Hatırlama (%)					
	TRAIN vs DEV			TRAIN+DEV vs TEST		
	SL	NSL	UA	SL	NSL	UA
Derinlikli Öğrenme ile Duygu Tespiti [27]	NA	NA	52.5	NA	NA	NA
Tez Çalışması (Derinlikli Öğrenme)	60	53	56	58	90	74
ITU MSPPR Lab. (SVM) [17]	89.1	97.2	93.2	79.9	80.1	80.0
Interspeech 2011 Kazanan (Hibrid) [28]	60.3	75.7	68.0	64.2	79.1	71.7
Interspeech 2011 Taban Seviye (SVM) [7]	NA	NA	67.3	NA	NA	70.3

1.3 Hipotez

Psikoakustik maskeleme yapılmış bürünsel öznitelikler, "Kısıtlandırılmış Boltmann Makinalar"ı kullanılarak ön eğitimi gerçekleştirilmiş bir sinir ağında, geri-yayma algoritması ile iyileştirilmiş ayrıştırıcı öğrenme işlevinden geçirilerek iki sınıf arasındaki tutarlılık bozulmayacak şekilde sınıflandırılabilir, yargısı tez çalışmasının üzerine kurulduğu hipotezdir ve bu hipotez, yapılan testler, ortaya konan sonuçlar ile kanıtlanmaya çalışılmıştır.

2. ÖZNİTELİK ÇIKARILMASI

2.1 Amaç

Ses verisinden durum sezimi, son yıllarda insan-makine etkileşimli uygulamalarda sıklıkla kullanılan "etkili bilgi işleme" metotları araştırma alanında çözülmeye çalışılan popüler bir problemdir. Bu alandaki birçok çalışma insani durumların makinelere öğretilmesi amacıyla yapılmıştır. Çalışmaların birbirinden farkları seçilen özniteliklerin ya da özniteliklere uygulanan istatistiki yaklaşımların farklılığı, kullanılan öznitelik sayısı ya da sınıflandırıcı farklılıklarından kaynaklanmaktadır.

Bu tez çalışmasında, konuşmadan uykulu/uykusuzluk durumlarını modellemeye uygun olduğu belirlenmiş bazı ses öznitelikleri ve sınıflandırma için derinlikli öğrenme algoritmaları kullanılarak "durum sezimi" problemi bir doğrusal olmayan sınıflandırma problemi olarak modellenmiş ve çözülmeye çalışılmıştır.

Literatürde kullanılan öznitelikler, algısal bürünsel özniteliklerdir. Anlamsal bağlamı olmayan bu öznitelikler, literatürde konuşmacı tanımadan duygu sezimine kadar değişik problemlerde yaygın olarak kullanılan temel ve perde frekanslar, enerji, konuşma hızı, MFCC(Mel-Frequency Cepstrum) katsayılarıdır. Fakat geleneksel öznitelikler insan kulağını tam olarak modelleyemediği için konuşmacı durum seziminden çok konuşmacı ve konuşma tanıma problemlerine daha uygundur. Çalışmada kullanılan özniteliklerin, yaygın olarak kullanılan bürünsel özniteliklerden farkı algısal bantta konuşmanın izgesel özelliklerini modelleyebilmesi ve aynı zamanda dile ait yapısal özelliklerle uğraşmadan içeriği zamansal bağlamda takip edebilmesidir [17]. Bu da sınıflandırmaya tabi tutulacak uykululuk düzeyine sahip konuşmanın izgesel ve zamansal boyutlarda modellemesini sağlamaktadır. Kullanılan bürünsel öznitelikler değişik seviyelerde yüksek seslilikte meydana gelen değişimleri yansıtacak şekilde formüle edilmiştir. Seslilikte meydana gelen değişimler hem kritik bantlar hem de zamansal değişimler üzerinden ölçülmektedir. Sesin algısal kalitesini ölçmek amacıyla kullanılan ITU [11] standardı, kullanılan özniteliklerin temelini oluşturmaktadır [4].Çalışmada kullanılan öznitelikler Çizelge

2.1'de listelenmektedir ve bu dokuz öznitelikten üçü Hertz frekans ölçeğinde, kalan altısı ise Bark frekans ölçeğinde hesaplanmaktadır.

Bu bölümde ilk olarak öznitelik çıkarma öncesi ses üzerinde uygulanan ön-işlemler anlatılmakta, ardından kullanılan özniteliklerin konuşma verisinden uykulu/uykusuz durumlarının sezilmesindeki rolleri açıklanmakta ve matematiksel formülasyonları verilmektedir.

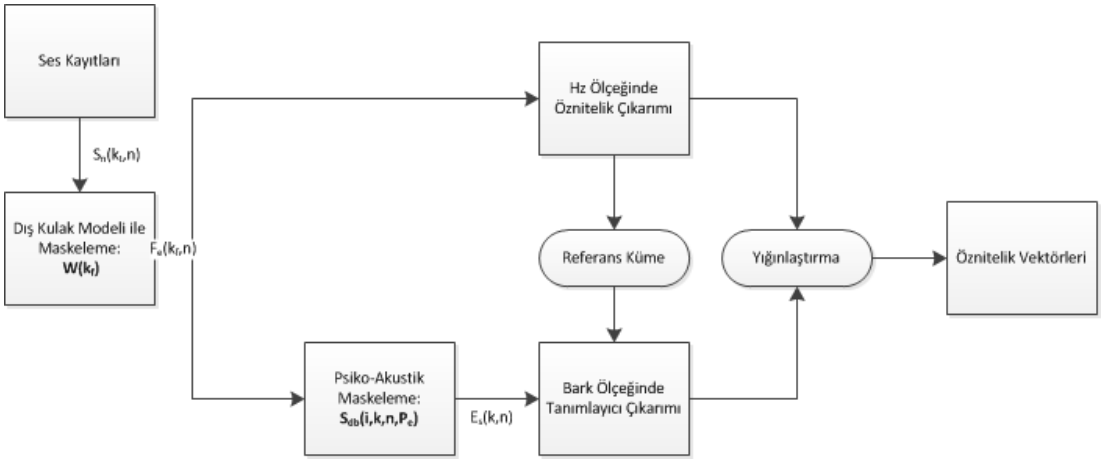
Çizelge 2.1 : Çalışmada kullanılan öznitelikler.

	Hz Ölçeğinde Hesaplanan Ses Öznitelikleri	
1	Ortalama Harmoniklik Genliği (OHG)	Ardışıl Y ses çerçevesi için kritik bantlardaki uykululuk değişimlerinin logaritmik izgesi ile kestirilmiş ortalama temel frekanslar
2	10db Algısal Bant Genişliği (W_{E1})	Gürültü eşiğini en az 10dB aşan en yüksek frekans bileşeni
3	5db Algısal Bant Genişliği (W_{E2})	Gürültü eşiğini en az 5dB aşan en yüksek frekans bileşeni
	Bark Ölçeğinde Hesaplanan Ses Öznitelikleri	
4	Normalize Edilmiş Uykululuk Düzeyi Farkı (NUF)	Uykulu/Uykusuz konuşmaların perde örnekleri ve ses çerçevesinin bark ölçeklemesiyle oluşturulmuş referans konuşma arasındaki maskelenmiş değişimlerin ortalaması
5	Normalize Edilmiş İzgesel Zarf Farkı (NIZD1)	Kritik banttaki ardışıl çerçevelerden alınmış referanstaki ölçülmemiş uykululuk perde örneklerinin normalize zarflarındaki değişim
6	Normalize Edilmiş İzgesel Zarf Farkı (NIZD2)	NIZD1'in tüm kritik bantlar üzerinden ortalaması
7	Normalize Edilmiş İzgesel Zarf Farkı (NIZD3)	NIZD1'in ardışıl Y ses çerçevesi üzerinden zamansal ortalaması
8	Ortalama Uykusuz Blok Sayısı (OUBS)	Belirli bir zaman aralığındaki uykusuz blokların beklenen sayısı
9	Çerçevenin Seslilik Değeri (ÇSD)	Dış kulak ağırlıklandırması uygulanan ses çerçevelerinin seslilik değerlerinin tüm kritik bantlar üzerinden toplamı

2.2 Ses Verisine Uygulanan Ön İşlemler

Öznitelik çıkarım aşamasından önce uygulanan ön işlemler Şekil 2.1'de görülen blok diyagramı üzerinden açıklanacaktır. Blok diyagramda görüldüğü gibi insan kulağının psiko-akustik ve algısal etkilerinin modellenebilmesi için ses verisinin bazı ön işlemlere tabi tutulması gerekmektedir. Bu işlemler temelde ITU tarafından önerilmiş

[11] algısal maskeleme adımlarının konuşmadaki algısal içeriğin belirginleştirilmesi için izgeye uygulanması ile gerçekleştirilir ve sonrasında öznitelik çıkarım aşamasına geçilir. Öznitelik çıkarım aşamasında, çalışmada kullanılacak düşük seviyeli öznitelikler Bark ölçeğindeki algısal izgeler ve Hertz ölçeğindeki akustik izgeler kullanılarak oluşturulur. En son aşamada ise elde edilen öznitelikler test ve eğitim aşamasında kullanılmak üzere gruplanır.



Şekil 2.1 : Öznitelik çıkarımı blok diyagramı, Sezgin, Gunsel, Kurt (2011)'den uyarlanmıştır.

Ön işlemlerde kullanılan işaretler 16 kHz'de örneklenmiş ve %50 örtüşme ile 43 ms'lik çerçevelere bölünmüş ses sinyalleridir. $s_n[k_t, n]$ ayrık ses örneğini göstermek üzere, n zaman-çerçevelerinin indeksi ve k_t zaman indeksidir. Her pencerelenmiş çerçeve için kısa zamanlı "Kısa Zamanlı Fourier Dönüşümü (STFT)" alınarak Fourier bölgesine geçilir ve denklem **(2.1)** elde edilir.

$$F[k_f, n] = \frac{1}{N_{FT}} \sum_{k_t=0}^{N_{FT}-1} h_w[k_t] s_n[k_t, n] e^{-\frac{j2\pi}{N_{FT}} k_f k_t} \qquad \textbf{(2.1)}$$

Fourier bölgesinde k_f frekans indeksi, $h_w[.]$ Hann pencereleme fonksiyonu olarak kullanılır. Çalışmada kullanılan Ayrık Fourirer Dönüşümü (N_{FT}) uzunluğu 2048'dir ve N_{FT} değişkeni ile ifade edilir.

İzgedeki algısal bileşenlerin belirginleştirilmesi için, ses sinyalinin ardışıl çerçeveleri ITU-PEAQ(International Telecommunications Union-Perceptual Evaluation of Audio Quality) [11] psiko-akustik modeli ile modellenmiştir. Bu kapsamda ilk önce orta ve dış kulak etkisini yansıtmak için izgesel bileşenler, frekans yanıtı orta ve dış kulak ağırlıklandırmayı modelleyen bir filtreden geçirilmiştir.

$$F_e[k_f, n] = |F[k_f, n]|.\, 10^{\frac{W[k_f]}{20}} \quad \textbf{(2.2)}$$

(2.2) ve **(2.3)** formülünde $W[k_f]$ ile gösterilen ağırlık fonksiyonu orta ve dış kulak etkisini yansıtan frekans cevabıdır.

$$W[k_f] = -0.6 \cdot 3.64 \cdot k_f^{-0.8} + 6.5 \cdot e^{-0.6 \cdot (k_f - 3.3)^2} - 10^{-3} \cdot (k_f)^{3.6} \quad \textbf{(2.3)}$$

Belirtildiği üzere 3 öznitelik Hertz ölçeğinde hesaplanırken, 6 öznitelik Bark ölçeğinde elde edilmektedir. Hertz ölçeğinden Bark ölçeğine dönüşüm **(2.4)** denklemi ile gerçekleştirilmiştir ve elde edilen Bark ölçeğindeki frekans bantları 1 kHz altında neredeyse doğrusal iken 1 kHz üzerinde üstel büyüyerek algısal bir filtre bankası oluşturur [11]. Bark ölçeği insan duyma sisteminin hassas olduğu 24 kritik bantı içerir. Böylelikle kulağın hassas olduğu frekansları daha iyi temsil eder.

$$z(Bark) \approx 7 \cdot \operatorname{arsinh}\left(\frac{f}{650}\right) \quad \textbf{(2.4)}$$

Bark bölgesindeki izgesel maskeleme etkisini modelleyebilmek için, enerji bileşenleri yayıcı fonksiyon, $S_{dB}(.)$, ile konvolüsyon işlemine tabi tutulur. Bark dönüşümüyle denklem **(2.5)**'teki $P_e[k, n]$ gösterimini alan dış kulak ağırlıklandırılmış enerji fonksiyonu yayılma fonksiyonunun derecesine bağlı olarak komşu frekans bantlarını maskeler. k frekans bandındaki *i* enerji bileşeni için yayılma fonksiyonu, f_c Bark ölçeğinin merkezi olmak üzere, denklem **(2.5)** ile gösterilen iki yönlü üstel bir fonksiyon olur.

$$S(i, k, n, P_e) = \begin{cases} 27(i-k)\Delta z & ; i \le k \\ \left[-24 - (230/f_c) + 2\log_{10}\left(P_e[k, n]\right)\right](i-k)\Delta z & ; i > k \end{cases} \quad \textbf{(2.5)}$$

$$\nabla z = i - k = 1/4$$

Denklem **(2.6)**'daki $E_s[k, n]$ yayılma fonksiyonu ve frekans cevabı kullanılarak her *n* ses çerçevesindeki her *k* kritik bandı için izgesel enerjinin frekanslara dağıtılması ile seslilik hesaplanır. Verilen **(2.6)** formülasyonundaki $B_s[k, n]$, 0 dB referans noktasında hesaplanmış normalizasyon katsayısıdır, N_c ise kritik bant sayısını gösterir ve değeri 109'dur.

$$E_S[k,n] = \frac{\left(\sum_{i=0}^{N_c-1} P_e[k,n] S_{db}(i,k,n,P_e[k,n]^{0.4})\right)^{\frac{1}{0.4}}}{B_S[k,n]} \quad (2.6)$$

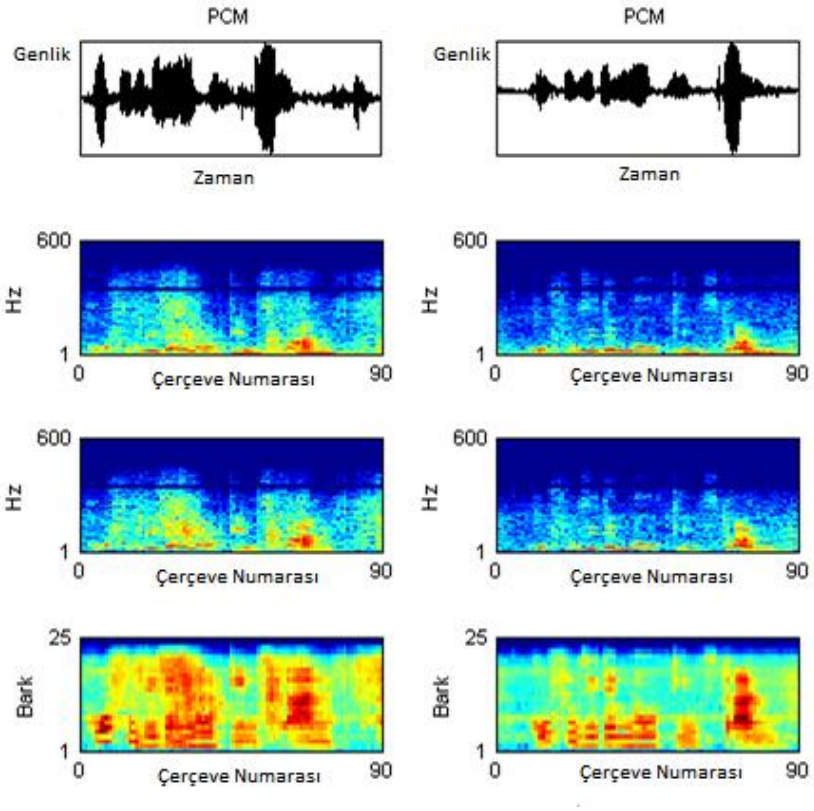
İzgesel maskeleme etkisi modellendikten sonra ileriye doğru maskeleme etkileri için kullanılacak zaman bölgesindeki yayılmalar modele katılır. İleriye doğru maskelemeyi modellemek için her bir kritik banttaki enerji seviyeleri zamanla denklem **(2.8)**'deki gibi yayılırlar. Denklemde kullanılan "*a"* sabiti, her kritik bandın merkez frekansı ile değişen bir zaman sabitidir.

$$\bar{E}[k,n] = a \cdot \bar{E}[k,n-1] + (1-a) \cdot E_s[k,n]^{0.3} \quad (2.7)$$

$$E[k,n] = \max\left(\bar{E}(k,n), E_s(k,n)\right) \quad (2.8)$$

$$\bar{E}[k,0] = 0$$

Algısal öznitelikler yukarıda tanımlanan seslilik düzeyleri kullanılarak hesaplanırlar. Şekil 2.2'de görülebileceği gibi NSL etiketli çerçeveler sol, SL etiketli çerçeveler sağ sütun olmak üzere, yukarıdan aşağıya ilk satır zaman bölgesinde kayıtlar, ikinci satır güç izgeleri, üçüncü satır dış kulak ağırlıkları ile ağırlıklandırılmış güç izgeleri, ve son satır bark ölçeğinde maskeleme yapılmış izgeleri göstermektedir. SL etiketli kayıtlar düşük frekanslarda NSL etiketli kayıtlara göre daha çok bilgi taşımaktadır. Maskeleme sonucunda elde edilen çizimlerde görülebileceği gibi, maskeleme zaman ve frekans bölgesinde daha iyi ayırt edilebilirlik sunmaktadır. Algısal maskelemenin avantajı frekans ve zaman bölgesindeki geçişlerin insan kulağına uygun olarak gözlemlenebilmesini mümkün kılmasıdır.



Şekil 2.2 : SL(1.Sütun)-NSL(2.sütun) sınıflarındaki konuşma kayıtlarından elde edilen ön işlem sonuçları. (a) Zaman bölgesindeki çizimler (b) Güç izgeleri (c) Dış kulak ağırlıkları ile ağırlıklandırılmış güz izgeleri (d) Bark ölçeğinde maskeleme yapılmış izgeler.

2.3 Ses Özniteliklerinin Çıkarılması

2.3.1 Normalize Edilmiş Uykululuk Düzeyi Farkı (NUF)

Konuşma sesliliğinin kritik bantlardaki değişimleri uykululuk-uykusuzluk ayırt etme probleminde kullanılan ayırt edici özniteliklerden biridir. Belirli bir referansa göre elde edilen değişimler kullanılacağı için, algılanan seslilikteki değişimler olarak ifade edilir. Çalışmada referansların her zaman diğer sınıftan seçilmesinden başka, referans seçiminde başka bir koşul ya da kural kullanılmamıştır.

Daha önceki alt bölümde formüle edilen, *n*. konuşma verisi için *k.* kritik bantta hesaplanmış seslilik değeri, dış kulak ağırlıklandırılmış enerjidir, $P_{e_{SL/NSL}}[k,n]$. *n.* ses çerçevesinde *k.* kritik bant, için hesaplanmış dış kulak ağırlıklandırılmış enerjidir, $P_{e_R}[k,n]$, kullanılarak ardışıl *Y* çerçeve için *"Normalize Edilmiş Uykululuk Düzeyi Farkı"* **(2.9)** denklemiyle hesaplanabilir.

$$NUF = 10log_{10}\frac{1}{Y}\sum_{n=1}^{Y}\left(\frac{1}{Z}\sum_{k=0}^{Z-1}\frac{Pe_{SL/NSL}[k,n]-Pe_R[k,n]}{M[k,n]}\right) \quad (2.9)$$

Z, kritik bantların sayısı olmak üzere değeri 109'dur ve *n* çerçeve numarası olarak kullanılır. Uykulu ve uykusuz konuşma verileri arasındaki farkı daha görünür kılmak için, yerel olarak algısal seslilik yaratan alçak frekans bileşenleri uyarlamalı olarak maskelenmiştir. Maske, $M[k,n]$, 2kHz'e kadar olan frekansları doğrusal olarak filtreleyerek uyarlamalı yüksek geçiren filtre gibi çalışır.

2.3.2 Normalize Edilmiş İzgesel Zarf Farkı (NIZD1, NIZD2, NIZD3)

Farklı uykululuk düzeylerindeki zamana bağlı bileşenlerin değişimini gözlemlemek için, ardışıl konuşma pencerelerindeki kritik bantlar kullanılarak yayılmamış uyarım seviyelerindeki değişimler kullanılır, değişimler izgesel bölgede normalize edilmiş bir zarf oluşturur ve bu zarf konuşmacının uykululuk düzeyi hakkında bilgi verir.

k. kritik banttaki *n.* çerçeve için elde edilmiş yayılmamış uyarım sesliliği, $E_S[k,n]$, izgesel zarfındaki değişimler, $E_{der}[k,n]$, **(2.10)** denklemi ile bulunabilir. Formüldeki *a* parametresi *0<a<1* arasında olup geçmiş çerçevelerin işlem gören ses çerçevesine olan etkilerini yansıtarak, sönümleme parametresi gibi davranır. Yapılan gözlemler sonucunda zarftaki değişimlerin uykusuz etiketli konuşmalarda, uykululara göre daha çok olduğu görülmüştür [17].

$$\bar{E}_{der}[k,n] = a\cdot\bar{E}_{der}[k,n-1] + (1-a)\cdot\left|E_S[k,n]^{0.3} - E_S[k,n-1]^{0.3}\right| \quad (2.10)$$

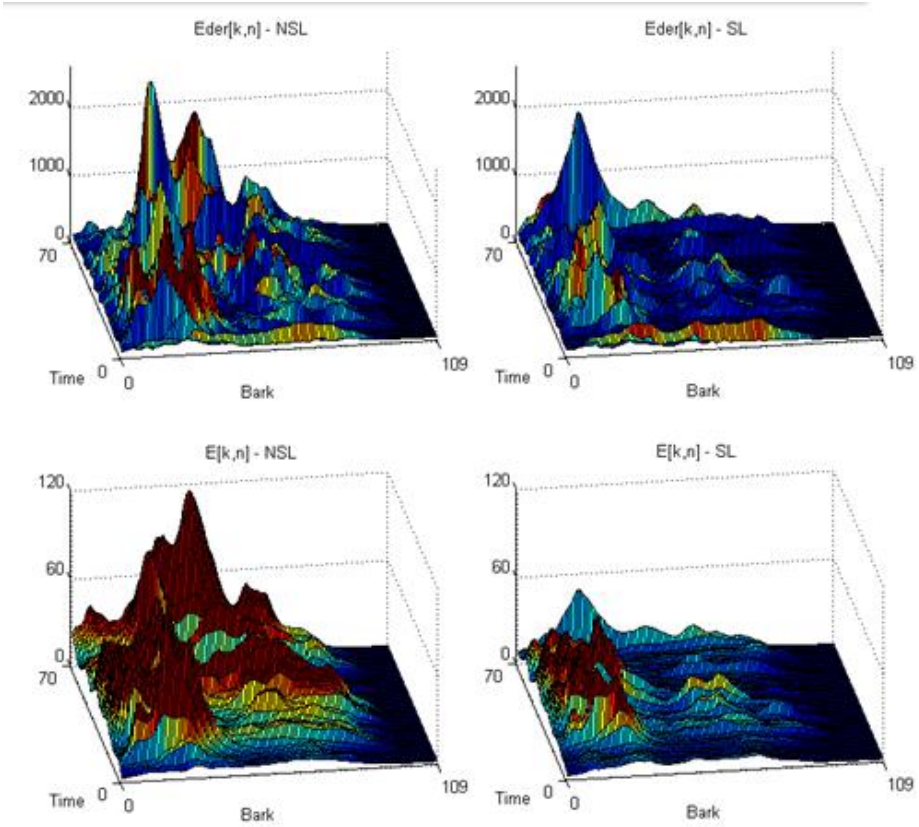
$$NIZ[k,n] = \frac{\bar{E}_{der}[k,n]}{1+(\bar{E}[k,n]/0.3)} \quad (2.11)$$

Normalize izgesel zarf diğer bir deyişle ardışıl çerçeveler arasındaki uykululuk değişimlerinin oluşturduğu seslilik demektir. **(2.11)** formülünün paydasında yer alan **(2.7)** denklemimde elde edilmiş $\bar{E}[k,n]$ terimi zamanla enerji sevilerinde yaşanan yayılmayı göstermektedir ve adaptif normalizasyon terimi görevini görmektedir.

NUF özniteliğinin hesaplanışına benzer şekilde, ayrımı güçlendirebilmek için uykulu ve uykusuz olarak etiketlenmiş konuşmaların normalize edilmiş izgesel zarfları(NIZ) düzeyleri belirli referans ile karşılaştırılır ve "*Normalize Edilmiş İzgesel Zarf Farkı (NIZD1)*" özniteliği elde edilmiş olur. β terimi $NIZ_{SL/NSL}[k,n]$ ve $NIZ_R[k,n]$ arasındaki minimum farkı kontrol etmek için kullanılır.

$$NIZD1[k,n] = w.\frac{|\,NIZ_{SL/NSL}[k,n] - NIZ_R[k,n]\,|}{\beta + NIZ_{SL/NSL}[k,n]} \qquad \textbf{(2.12)}$$

Denklem **(2.12)** ile elde edilen öznitelikten istatistiksel yöntemler kullanılarak 2 istatistiksel tanımlayıcı daha elde edilir. İlk istatistiksel tanımlayıcı(NIZD2), normalize edilmiş farkların N_c=109 Bark ölçeği üzerinden ortalamasının alınmasıyla elde edilir, ikinci istatistiksel tanımlayıcı (3. Öznitelik, NIZD3) ise ardışıl Y pencereden hesaplanan normalize edilmiş farkların zamansal ortalamasının alınmasıyla elde edilir. Şekil 2.3'te görüldüğü gibi zarfta yaşanan değişiklikler NSL etiketli kayıtlarda SL etiketli kayıtlara göre daha fazladır.



Şekil 2.3 : NSL ve SL etiketli kayıtlar arasındaki zarf değişimlerinin farkı özniteliğinin kritik bantlarda zamana bağlı değişimi.

2.3.3 Ortalama Uykusuz Blok Sayısı (OUBS)

"Ortalama Uykusuz Blok Sayısı" özniteliği Bark ölçeğinde elde edilen ardışıl çerçeve gruplarında (bu çalışmada ardışıl 70 çerçeve kullanılmıştır) gözlenen yüksek seslilik seviyelerinin bir ölçüsüdür.

k. Bark ölçeğindeki *n.* ses çerçeveleri *"k"* ve *"n"* ile gösterilmek üzere **(2.13)** denklemindeki $e[k,n]$ referans $E_{sR}[k,n]$ ve gözlenen $E_{s(SL/NSL)}[k,n]$ konuşma kayıtları arasındaki seslilik seviyeleri farkını gösterir.

$$e[k,n] = 10\log_{10}\frac{E_{s(SL/NSL)}[k,n]}{E_{sR}[k,n]} \qquad \textbf{(2.13)}$$

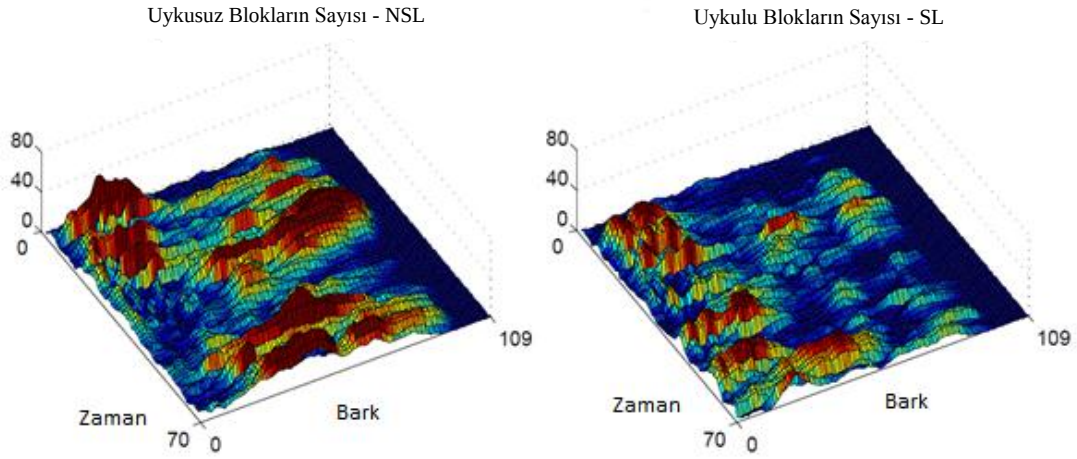
Belli bir zaman ağırlığındaki uykusuz etiketli blokların beklenen sayısını hesaplayabilmek için, **(2.14)** ile ortaya konan olasılıksal bir yaklaşım kullanılır, bu yaklaşımla belirli bir seslilik eşik değerinin üzerinde kalan ses pencerelerinin sayısı kestirilmeye çalışılır [17]. Yayılma fonksiyonunun olasılığının, $p[k,n]$, hesaplandığı denklem **(2.14)**'te $s[k,n]$ normalizasyon katsayısı iken *k* sabit bir sayıdır.

$$p[k,n] = 1 - \left(\frac{1}{2}\right)^{\left(\frac{e[k,n]}{s[k,n]}\right)^b} \qquad \textbf{(2.14)}$$

Gözlemlenen çerçevelerin ilintisiz olduğu kabul edilirse, *n*. konuşma çerçevesinin uykusuz olarak etiketlenme olasılığı deklem **(2.15)** ile hesaplanabilir.

$$P[n] = \prod_{\forall k} p[k,n] \qquad \textbf{(2.15)}$$

Doğası gereği konuşmacının uykusuz olduğu durumlarda yapılan ses kayıtlarında daha yüksek seslilik seviyeleri gözlendiği için ayırt edicilik açısından önemli bir özniteliktir. NSL etiketli kayıtlarda beklenen sık yaşanan seslilik değişimleri, Şekil 2.4'te de görülebileceği gibi, NSL-SL için iyi bir ayırt edicidir.



Şekil 2.4 : NSL ve SL etiketli kayıtlarda uykusuz blokların sayısının kritik bantlarda zamana bağlı değişimi.

2.3.4 Çerçevenin Seslilik Değeri (ÇSD)

Bir ses çerçevesinin kritik banttaki seslilik değeri, $L[k,n]$, **(2.16)** ve **(2.17)** denklemi ile hesaplanır ve $E_{IG}[k]$ iç gürültü olarak hesaplanır. $E[k,n]$ daha önceki bölümlerde de anlatıldığı gibi, *n.* ses çerçevesinin *k.* kritik bandındaki seslilik değeridir.

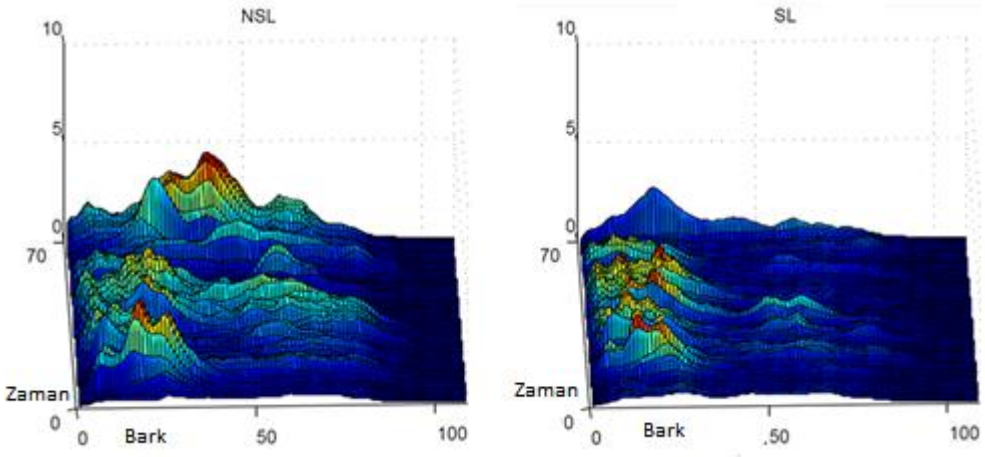
$$L[k,n] = const\,.\left(\frac{1}{s[k]}\frac{E_{IG}[k]}{10^4}\right)^{0.23}\left[\left(1 - s[k] + \frac{s[k]\,E[k,n]}{E_{IG}[k]}\right)^{0.23} - 1\right] \qquad \textbf{(2.16)}$$

$$s[k] = 10^{0.1(-2-2.05\,atn\left(\frac{f}{4000}\right)-0.75\,atn\left(\left(\frac{f}{1600}\right)^2\right))} \qquad \textbf{(2.17)}$$

Bir çerçevedeki kritik bantlar üzerinden hesaplanan tüm seslilik değerleri (ÇSD) ise, tüm filtre kanallarındaki (Z) seslilik değerlerinin toplanması ile **(2.18)**'deki gibi elde edilir.

$$\boldsymbol{L_{tüm}[n] = \frac{24}{Z}\sum_{k=0}^{Z-1} max(L[k,n], 0)} \qquad \textbf{(2.18)}$$

Şekil 2.5'te görülebileceği gibi SL etiketli kayıtların aksine NSL etiketli kayıtlarda seslilik kritik bantların neredeyse hepsini bastırmaktadır.



Şekil 2.5 : NSL ve SL etiketli kayıtlarda seslilik düzeyi farklarının kritik bantlarda zamana bağlı değişimi.

2.3.5 Ortalama Harmoniklik Genliği (OHG)

Temel frekanstaki değişimleri yansıtan ve konuşmanın harmonik yapısı ile ilgili bir özniteliktir. Doğrusal olmayan(ör. Bark ölçeği) frekans dönüşümleri harmonik yapıyı yayarak iki sınıf arasındaki ayrımı zorlaştırdığı için, özniteliğin elde edilebilmesi için Hz ölçeğinde doğrusal frekans izgeleri kullanılacaktır. Öznitelik çıkarılırken ilk adımda, referans ve gözlenen ses sinyali için dış kulak ağırlıklandırılmış enerji

farkları kritik bantlar üzerinden hesaplanır ve uykululuk düzeyi ile kritik bantlar arasındaki ilinti elde edilmiş olur. Ardışıl *Y* konuşma çerçevesi için kestirilen temel frekansların ortalaması OHG özniteliği olarak kullanılır. Geleneksel yöntemlerde temel frekanslar ilinti fonksiyonunun logaritmik izgesi kullanılarak kestirilir, bu çalışmada kritik bantlar ile uykululuk düzeyi farkları ilintilendirilmiş olup zaman bölgesindeki konuşma sinyali kullanılmamıştır.

İzgesel indeksi k_f olan *n.* çerçeve için referans küme ile farkı gösteren fonksiyon, **(2.19)**'da gösterilen $P_{e_{Fark}}[k_f, n]$, olmak üzere frekans bölgesinde gözlenen, $F_{eG}[k_f, n]$, ve referans konuşmalar, $F_{eR}[k_f, n]$, için izgesel enerji kullanılarak bulunabilir.

$$P_{e_{Fark}}[k_f, n] = log\left(\left|F_{eG}[k_f, n]\right|^2\right) - log\left(\left|F_{eR}[k_f, n]\right|^2\right) = log\frac{\left|F_{eG}[k_f, n]\right|^2}{\left|F_{eR}[k_f, n]\right|^2} = P_{k_f} \quad \textbf{(2.19)}$$

P_{k_f} satır vektörü, 256 ardışıl çerçeve için k_f kritik bantlarındaki enerji farklarının gruplanmasıyla oluşturulur. Denklem **(2.20)**'ile bulunan 256 komşuluk üzerinden normalize edilmiş otokorelasyon fonksiyonu, $C[l]$, komşular arasındaki uykululuk ve uykusuzluk düzeyleri arasındaki farkı göstermektedir.

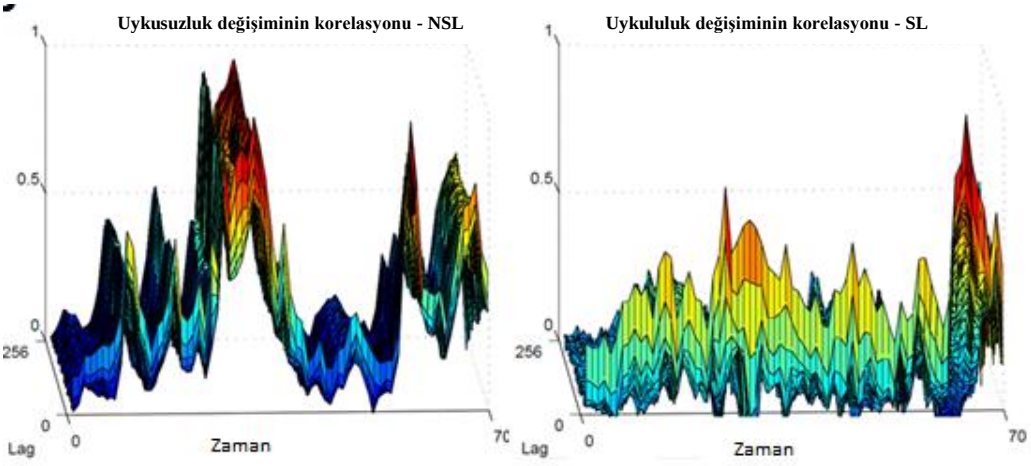
$$C[l] = \frac{P_{k_f} P_{k_f+l}}{\sqrt{\left|P_{k_f}\right|^2 \left|P_{k_f+l}\right|^2}} \quad \textbf{(2.20)}$$

$$S[k_f] = \left|\frac{1}{256}\sum_{i=1}^{256} C[l] e^{j2\pi k_f l/256}\right|^2 \quad \textbf{(2.21)}$$

Normalize edilmiş otokorelasyon fonksiyonu kullanılarak denklem **(2.21)**'de hesaplanan güç izgesinin tepe değeri harmonik genliği, **(2.22)**, $E_{Hmax}[n]$, verir. OHG özniteliği de *Y* ardışıl çerçevenin genliklerinin ortalaması alınarak hesaplanır.

$$OHG = \frac{1000}{Y}\sum_{n=0}^{Y-1} E_{Hmax}[n] \quad \textbf{(2.22)}$$

Şekil 2.6'da görülebileceği gibi uykusuzluk taşıyan konuşmaların, tonlamada artış ve azalışlara sahip uykululuk taşıyan konuşmalarla karşılaştırınca daha dengeli harmonikleri ile periyodik sinyale benzediği görülür.



Şekil 2.6 : NSL ve SL etiketli kayıtlar için uykululuk değişiminin kritik bantlar korelasyonunun zamana bağlı değişimi.

2.3.6 Algısal Bant Genişliği

Algılanan ses rengi, donukluk, boğukluk gibi etkiler, uykululuk/uykusuzluk durumunda kaydedilen konuşmalar farklı algısal bant genişliklerinde değişiklik gösterirler. Böylelikle bu iki sınıfı ayırmak için sinyal bant genişliğindeki değişimler kullanılır. Özniteliğin algısal olarak değerlendirilmesinin nedeni, maskelenmiş izge kullanılarak hesaplanmasıdır.

Özniteliğin hesaplanışında, üst frekans aralığındaki frekans izgesinin maksimum değeri, **(2.23)**'teki $W_1[n]$, gürültü tabanı olarak kullanılır ve gürültü tabanını alçak frekanslara doğru 10 dB ve 5 dB aşan yüksek frekans bileşenleri kestirilmiş algısal bant genişliği olarak kullanılırlar. Referans ses sinyalinde gürültü tabanını 10 dB aşan ilk frekans bileşeni *n.* çerçeve için bant genişliği olur, **(2.24)**'teki $W_2[n]$, gürültü tabanının 5 dB aşan ilk frekans bileşeni, denklem **(2.25)**, ise ikinci bant genişliği olarak kullanılır ve *"Algısal Bant Genişliği"* kullanılarak hesaplanan 2. öznitelik de elde edilmiş olur.

$$W_1[n] = \max_{k_f}\{F[k_f, n]\}, 14.4kHz < k_f < 16kHz \quad \textbf{(2.23)}$$

$$W_2[n] = \arg_{k_f}\{F[k_f, n] > 10log(F[W_1[n], n])\}, 3kHz < k_f < 14.4kHz \quad \textbf{(2.24)}$$

$$W_3[n] = \arg_{k_f}\{F[k_f, n] > 5log(F[W_2[n], n])\}, \{F[k_f, n] < 5log(F[W_2[n], n])\}, k_f < W_2[n] \quad \textbf{(2.25)}$$

Her çerçeve için hesaplanan bu kestirim değerlerinin denklemler **(2.26)** ve **(2.27)**'deki gibi Y ardışıl çerçeve üzerinden ortalaması alınarak elde edilen değerler, öznitelik olarak kullanılır.

$$W_E = \frac{1}{N}\sum_{n=0}^{Y-1} W_2[n] \tag{2.26}$$

$$W_{E1} = \frac{1}{N}\sum_{n=0}^{Y-1} W_3[n] \tag{2.27}$$

2.4 Sözlük Öğrenmesi

İki sınıflı bir öğrenme problemine dönüştürülmüş konuşmacı uykululuk/uykusuzluk durum tespiti problemi aslında 10 sınıftan oluşan bir verinin iki ana sınıfa sınıflandırılması problemidir ve saçılmaların iyi yönetilmesi gerekir. Saçılmaları iyi yönetebilmek için önerilen çözüm ise uykululuk seviyeleri için birer sözcük kümesi yani öbekler oluşturmaktır [17]. Çalışmada öbeklerin elde edilebilmesi için algısal özniteliklere *k-ortalamalı öbekleme* uygulanmış olup *k* burada eğitim ve testte kullanılacak kod kelimelerinin sayısını göstermektedir.

Teorik olarak, sınıflandırmada öbeklenmiş veri kullanılacağı durumda sınıflandırıcı performansını belirleyecek ana etken öbekleme sonucunda ortaya çıkan kod sözcüğü sayısı olacaktır ve alt sınıf sayısı 10 olduğu için en az 10 öbek bulunması gerekecektir. Optimum sayının bulunması için çalışmada öbeklere dahiliyeti belirleyecek olan siluetler olacaktır. Ana eğitim kümesinde kullanılan vektör sayısı *L*, *t.* iterasyonunda x_i özniteliği için hesaplanan siluet değeri $s^t(i)$ ile gösterilir ve **(2.28)** denklemi ile siluet değeri hesaplanır. Denklemde yer alan $a^t(i)$ değeri x_i vektörünün öbekte yer alan diğer vektörler ile ortalama farklılığını ifade eder, $b^t(i)$ değeri ise x_i vektörünün diğer tüm öbekler de dahil en düşük ortalama farklılığı ifade eder. Siluet değeri 1 ise x_i vektörü öbek ile tam benzerlik, 0'a yakın ise de 2 öbeğe de aynı benzerliği, -1'e yakın ise de vektörün yanlış öbekte yer aldığını göstermektedir. Vektörün yanlış öbekte yer aldığı durum için yeniden öbekleme işlemi gerçekleştirilir. Ortalama siluet değeri ise *L* vektör için ortalama bulunarak hesaplanır.

$$s^t(i) = \begin{cases} Eğer\ a^t(i) < b^t(i), & 1 - a^t(i)/b^t(i), \\ Eğer\ a^t(i) = b^t(i), & 0, \\ Eğer\ a^t(i) > b^t(i), & b^t(i)/a^t(i) - 1 \end{cases} \tag{2.28}$$

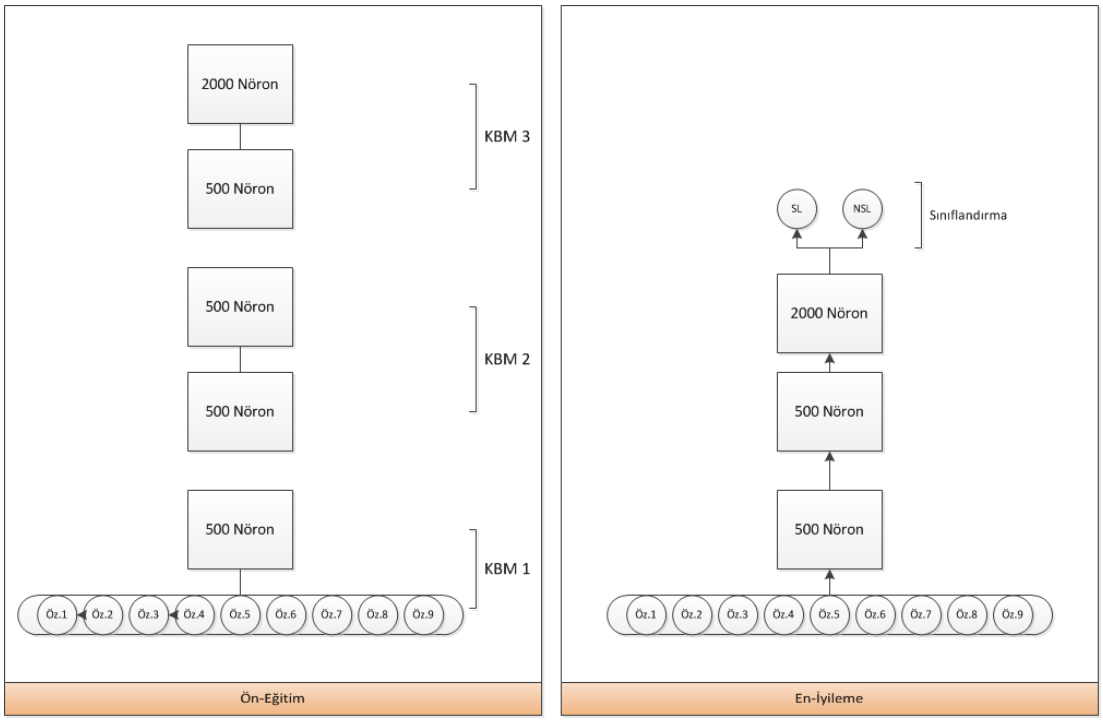
Sözlük öğrenme algoritmasında büyük bir *k* değerinden başlanarak, uygun siluet değerine varana kadar *k-ortalamalı öbekleme* tekrarlanır ve k azaltılarak en iyi değerine ulaşılır. Çalışmada kullanılan öznitelikler konuşmaların tamamı yerine çerçeveler üzerinden hesaplandığı için öbeklemeden önce yığınlaştırma işlemi gerçekleştirilir. Bu işlemde çerçeve bazlı hesaplanan başlangıç vektörlerinin *Y* ardışıl vektör üzerinden ortalaması alınarak, Y ardışıl çerçeveyi temsil eden bir ortalama vektör oluşturulur. Bu çalışmada kullanılan *Y* değeri 70 olup, öznitelik hesaplama işlemi %50 örtüşen 43 ms'lik çerçeveler üzerinden hesaplanır, bu da ortalama yığın uzunluğunu 1 sn yapar.

3. DERİNLİKLİ ÖĞRENME İLE KONUŞMACI DURUM MODELLEME

Derinlikli öğrenme, makine öğrenme algoritmaları kullanılarak doğrusal olmayan dönüşümlerle öğrenilecek verinin en iyi temsilini bulmayı amaçlayan bir yaklaşımdır.

Derinlikli öğrenme yaklaşımının temelini yapay sinir ağları oluşturur ve derinlikli öğrenme yapay sinir ağlarını geliştirerek beynin çalışmasını daha iyi modelleyen yöntemler önerir. Yapay sinir ağlarının performansını arttırmayı hedefleyen bu teori, yoğun matris işlemleri gerektirdiğinden, donanım kısıtları nedeniyle 1980'lerde uygulamaya dönüşememiştir. 80'lerin sonlarında Hinton ve LeCun'un önerdiği geri yayma algoritması [29], bilim çevrelerinde büyük yankı uyandırmasa da Kanada İleri Araştırmalar Enstitüsü'nün dikkatini çekmiştir ve belki de bu adım sayesinde Montreal ve Toronto Üniversitelerindeki makine öğrenmesi grupları derinlikli öğrenme konusunda referans alınan gruplar olarak öne çıkmıştır.

Derinlikli öğrenme algoritmaları kullanılarak farklı yapılar kurulabilmektedir, bu çalışmada konuşmacının SL/NSL durumları iki aşamalı bir eğitim modeli ile öğrenilmiş ve sınıflandırıcı eğitimi tamamlanmıştır. Bu kapsamda eğitim ikiye ayrılarak ön-eğitim ve eniyileme alt adımları oluşturulmuştur ve Şekil 3.1'de görülen blok diyagramlar ile modellenen bir eğitim algoritması gerçeklenmiştir. Ön-eğitim adımında üretici (generative) öğrenme ile parametrelere rasgele başlangıç değerleri verilerek etiketlenmemiş veri için model parametreleri ve bunları başlangıç koşulu alan ayrıştırıcı (discriminative) öğrenme adımında model parametreleri özyinelemeli olarak öğrenilir. Bölüm 3.1'de detaylandırılacak olan ön-eğitim modeli derin KBM ağlarından oluşur ve üç katmanlıdır. Bölüm 3.2'de anlatılacak eniyileme adımında ön-eğitim ile elde edilmiş parametrelerin etiketlenmiş veri kullanılarak eniyilemesi yapılır. Ayrıştırıcı öğrenme kullanılan bu basamakta geri-yayma algoritması ile istenen ve güncel çıkışlar arasındaki hata en küçüklenmeye çalışılır, ve en küçükleme için eşlenikli eğim algoritması kullanılır [15].



Şekil 3.1 : Derinlikli öğrenme blok diyagramı. (a) Ön eğitim modeli. (b) Eniyileme modeli.

3.1 Ön-Eğitim Modeli

Bu çalışmada, KBM kullanılarak yapılan bir öğrenmenin gerçeklenmesinde "Karşıtlıklı Iraksay" (Contrastive Divergence, "CD") [13] öğrenme prosedürü kullanılmıştır. Bu seçim; öğrenme hızı, ağırlık maliyeti, ağırlıkların başlangıç değerleri, kullanılan saklı katmandaki işlem birimi sayısı, işlenecek yığın kümelerinin büyüklükleri gibi önemli parametrelerin değerlerinin belirlenmesinin önemini beraberinde getirmektedir. Bu alandaki pratik tecrübeye dayanan bu seçimleri yaparken Geoffrey E. Hinton ve ekibinin çalışmalarındaki model referans alınarak çalışma kurulan bu altyapı üzerine konumlandırılmıştır [9].

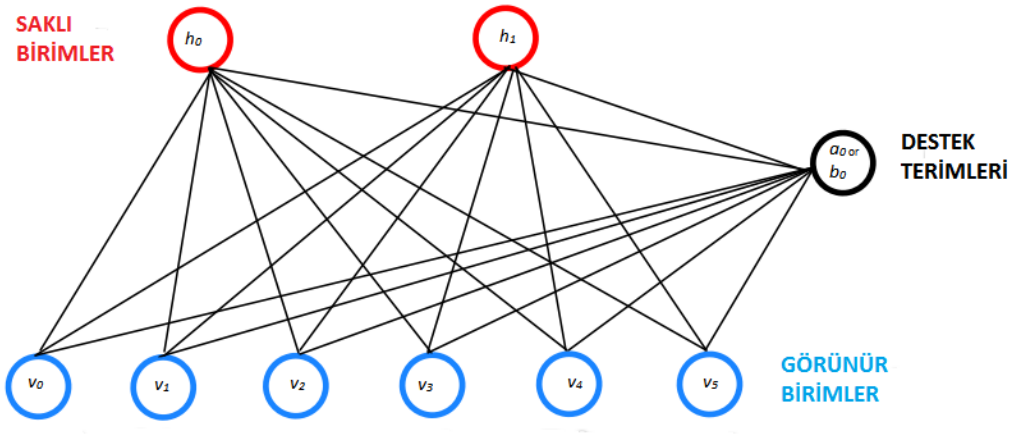
Kısıtlı Boltzmann Makinaları kullanmanın avantajları şu şekilde sıralanabilir:

- Etiketlenmemiş uykulu ve uykusuz konuşma kayıtlarının kullanılması ile öğrenme gerçekleşir, dolayısıyla oldukça yüklü ve subjektif bir iş olan etiketlenmiş veri kümesi oluşturulmasına ihtiyaç duymaz.
- Kurgulanan ağ anlamlı ve değerli nitelikleri kendisi öğrenir.

Bu nedenle çalışmamızda ön eğitim algoritması olarak KBM kullanılmıştır.

3.2 Kısıtlandırılmış Boltzmann Makinaları ve Karşıtlıklı Iraksay Algoritmaları

KBM iki ana katmandan oluşur. Birinci katman değerlerini gözlediğimiz "görünür katman", ikinci katman ise gözlenen değerlerin üretilmesine yol açan "saklı katman" olarak adlandırılır. Bu iki katman arasındaki işlem birimleri arasında yönsüz bağlantılar bulunur, fakat saklı katmanın kendi işlem birimleri arasında ve görünür katmanın kendi işlem birimleri arasında herhangi bir bağlantı bulunmamaktadır. Şekil 3.2'de standart bir KBM ağ yapısı görülmektedir.



Şekil 3.2 : Standart KBM ağ yapısı.

"Yönsüz Graf Tabanlı Modeller"e örnek olan KBM giriş verisi üzerinden (görünür katman) saklı katmanı da kullanarak hedef bir olasılık dağılımını öğrenir ve bu sayede gözlenen veriye uyan bir model oluşturmayı amaçlar. Bu dağılımın enerjinin üstel bir dağılımı olduğu varsayımına dayanılarak olasılığı enbüyükleyen çözümün bulunması tanımlanan enerji fonksiyonunun enküçüklemesi ile sağlanır. Standart KBM yapısında kullanılan enerji fonksiyonu görünür ve saklı katmanlara arası ilişkiyi modeller ve denklem (**3.1**)'de tanımlanmıştır.

$$E(v,h) = -\sum_{i \in \text{görünür}} a_i v_i \; - \sum_{j \in \text{saklı}} b_j h_j \; - \sum_{i,j} v_i h_j w_{ij} \qquad \textbf{(3.1)}$$

Denklem **(3.1)**'de görülen v_i ve h_j terimleri sırasıyla, *i.* görünür işlem birimi ve *j* . saklı işlem birimine ilişkin durumları ; a_i ve b_j destek (bias) terimlerini; w_{ij} ise katmanlar arasındaki ağırlık matrisinin ilgili ağırlık elemanını temsil etmektedir.

KBM ağ yapısı tüm görünür ve saklı işlem birimi eşleşmelerinden doğan vektör çiftlerine enerji fonksiyonu yardımıyla bir olasılık atar. Elde edilen bu olasılık

dağılımı denklem **(3.2)**'de görülen Gibbs olasılık dağılım fonksiyonu yardımıyla hesaplanabilir. Denklem **(3.2)**'de görülen payda terimi "Z" tüm olası ağırlıklar için aynı kalacağından (denklem **(3.3)**) hesaplanması gerekmeyen normalizasyon terimi olarak kullanılmaktadır.

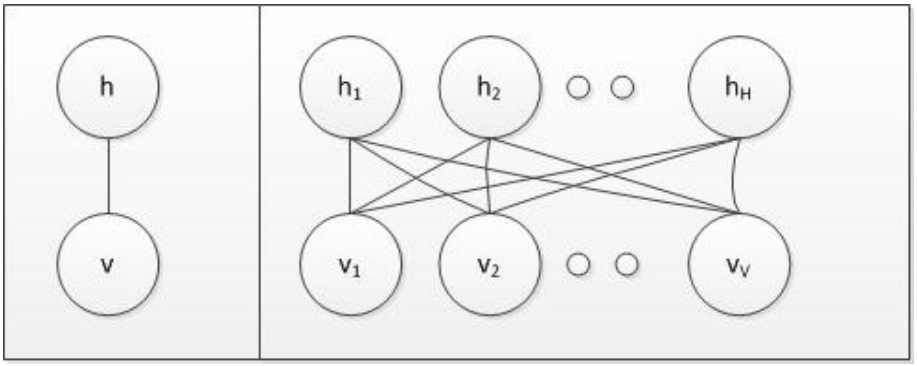
$$p(v,h) = \frac{e^{-E(v,h)}}{Z} = \frac{exp(-E(v,h))}{Z} = \frac{1}{Z}\exp(h^T wv)\exp(a^T v)\exp(b^T h) \qquad \textbf{(3.2)}$$

$$Z = \sum_{v,h} e^{-E(v,h)} \qquad \textbf{(3.3)}$$

Denklem **(3.2)**'deki olasılık dağılımı, her bir görünür ve saklı katman elemanının birbirinden bağımsız olduğu varsayımı ve görünür ve saklı katmanların birbirinden bağımsız olduğu varsayımı altında **(3.4)** denklemindeki gibi ifade edilir.

Buradan hareketle skaler olarak gösterilimler tekrarlanılarak da çarpanlar gösterilebilir. Denklem **(3.1)**'de görülen enerji eşitliği **(3.2)** denklemi ile birlikte kullanıldığında denklem **(3.4)** elde edilmektedir.

$$p(v,h) = \frac{1}{Z}\prod_j \prod_i \exp(v_i h_j w_{ij}) \prod_i a_i v_i \prod_j b_j h_j \qquad \textbf{(3.4)}$$



Şekil 3.3 : Vektörel ve Skaler Gösterilimler.

KBM ağının görünür bir "v" vektörüne atadığı olasılık değeri, tüm olası saklı işlem birimlerinin oluşturdukları vektörlerin, *h'*, toplamı ile denklem **(3.5)** kullanılarak bulunabilir.

$$p(v) = \sum_{h'} p(v,h') = \frac{1}{Z}\sum_h e^{-E(v,h)} \qquad \textbf{(3.5)}$$

Denklem **(3.4)** ve denklem **(3.5)** Bayes eşitliğinde yerleştirilirse görünür bir v vektörüne karşı düşen saklı h vektörü için hesaplanacak sonsal (posterior) dağılım denklem **(3.6)**'daki koşullu olasılık işlevi ile ifade edilebilir.

$$p(h|v) = \frac{p(v,h)}{p(v)} \quad \textbf{(3.6)}$$

Denklem **(3.6)**'daki ortak olasılık işlevi **(3.7)** denklemindeki açık haliyle denklem **(3.6)**'nın payını ve *"h"* saklı işlem biriminin olası değerleri (0 ve 1) üzerinden toplam ifadesi ile gösterilimi ise paydasını oluşturarak Denklem (3.8) elde edilir. Denklem **(3.8)**'de toplam ifadesinde üst indis($\{0,1\}^H$) olarak yer alan *"H"* saklı işlem birimlerinin sayısını ifade etmektedir.

$$p(v,h) = \frac{1}{Z}\exp(h^T wv + a^T v + b^T h) \quad \textbf{(3.7)}$$

$$p(h|v) = \frac{\frac{1}{Z}\exp(h^T wv + \cancel{a^T v} + b^T h)}{\sum_{h' \in \{0,1\}^H} \frac{1}{Z}\exp(h'^T wx + \cancel{a^T v} + b^T h')} \quad \textbf{(3.8)}$$

Üstel fonksiyonun içerisindeki saklı işlem birimi içermeyen ifadeler üzerinden sadeleştirmeler yapılır ve skaler olarak ifade edilirse sonsal dağılım denklem **(3.9)**'daki biçimde elde edilir. Denklemlerde v_s ile gösterilen eleman görünür işlem birimlerinden oluşturulan satır vektörüdür ve $w_{j.}$ ise katsayılar matrisinin ilgili satırından elde edilen vektördür.

$$p(h|v) = \frac{\exp(\sum_j h_j w_{j.} v_s + b_j h_j)}{\sum_{h'_1 \in \{0,1\}} \cdots \sum_{h'_H \in \{0,1\}} \exp(\sum_j h'_j w_j v_s + b_j h'_j)} \quad \textbf{(3.9a)}$$

$$p(h|v) = \frac{\prod_j \exp(h_j w_{j.} v_s + b_j h_j)}{\sum_{h'_1 \in \{0,1\}} \cdots \sum_{h'_H \in \{0,1\}} \prod_j \exp(h'_j w_{j.} v_s + b_j h'_j)} \quad \textbf{(3.9b)}$$

$$p(h|v) = \frac{\prod_j \exp(h_j w_{j.} v_s + b_j h_j)}{\left(\sum_{h'_1 \in \{0,1} \exp(h'_1 w_1 . v_{sv} + b_1 h'_1)\right) \cdots \left(\sum_{h'_H \in \{0,1\}} \exp(h'_H w_H . v_{sv} + b_H h'_H)\right)} \quad \textbf{(3.9c)}$$

$$p(h|v) = \frac{\prod_j \exp(h_j w_{j.} v_s + b_j h_j)}{\prod_j (\sum_{h'_j \in \{0,1\}} \exp(h'_j w_{j.} v_s + b_j h'_j))} \quad \textbf{(3.9)}$$

İkili sayılar olan *0 ve 1'i* değer olarak alabilen saklı işlem biriminin değerleri yerine yazılırsa denklem **(3.10)** elde edilir.

$$p(h|v) = \frac{\prod_j \exp(h_j w_{j.} v_s + b_j h_j)}{\prod_j (1 + \exp(b_j + w_{j.} v_s))} = \prod_j \frac{\exp(h_j w_{j.} v_s + b_j h_j)}{(1 + \exp(b_j + w_{j.} v_s)} = \prod_j p(h_j \mid v) \quad \textbf{(3.10)}$$

Saklı işlem birimimiz ikili değerler aldığı için karşı düşen rastlantı değişken Bernoulli dağılımına sahiptir, ve bu özellikten yararlanarak 1 değeri için koşullu olasılık ifadesi denklem **(3.11)**'deki gibi bulunabilir.

$$p(h_j = 1 \mid v) = \frac{\exp(b_j + w_j.v_s)}{1+\exp(b_j + w_j.v_s)} \quad \textbf{(3.11)}$$

Pay ve payda *"exp(b_j+w_yv)",* ifadesi ile çarpılırsa üstel terimler 1 değerine eşit olur ve denklem **(3.12)** elde edilir. Denklem **(3.13)**'de **(3.12)** denkleminin sadeleştirilmesini sağlayan sigmoid fonksiyonunun genel ifadesi verilmektedir.

$$p(h_j = 1 \mid v) = \frac{1}{1+\exp(-b_j - w_j.v_s)} = sigm(b_j + w_j.v_s) \quad \textbf{(3.12)}$$

$$sigm(x) = \sigma(x) = \frac{1}{(1+\exp(-x))} \quad \textbf{(3.13)}$$

KBM yönsüz bir graf tabanlı model olduğu için görünür işlem birimleri için de benzer formülasyonlar yapılabilir. Benzer notasyon ile ifade edilen sonsal dağılım denklem **(3.14)**'te ve karşı düşen açık gösterim denklem **(3.15)**'te görülmektedir. Denklem **(3.15)**'de kullanılan $W_{.k}$ ifadesi ağırlık matrisinin *k.* sütununu ifade etmektedir.

$$p(v \mid h) = \prod_i p(v_i \mid h) \quad \textbf{(3.14)}$$

$$p(v_i = 1 \mid h) = 1/1 + \exp(-a_i - h^T w_{.k}) = sigm(a_i + h^T w_{.k}) \quad \textbf{(3.15)}$$

Denklem **(3.5)**'de tanımlanan önsel (prior) dağılım ifadesi ayrıntılı incelenirse KBM ağının "v" giriş vektörüne atadığı önsel olasılıklardan KBM'nin giriş dağılımlarını nasıl modellediği anlaşılabilir. Denklem **(3.5)** aşağıdaki adımlarda daha açık ifade edilirse denklem **(3.23)** genel gösterimi elde edilebilir. Denklem **(3.5)** olası *h* değerleri üzerinden ifade edilirse denklem **(3.16)** ve denklem **(3.16)**'daki enerji fonksiyonunun açık ifadesi yazılırsa **(3.17)** denklemi elde edilir.

$$p(v) = \sum_{h\in\{0,1\}^H} p(v,h) = \frac{1}{Z}\sum_{h\in\{0,1\}^H} e^{-E(v,h)} \quad \textbf{(3.16)}$$

$$p(v) = \sum_{h\in\{0,1\}^H} \frac{1}{Z}\exp(h^T wx + a^T v + b^T h) \quad \textbf{(3.17)}$$

Denklem **(3.17)** tüm (*H*) saklı işlem birimleri üzerinden skaler terimler kullanılarak ifade edilerek Denklem **(3.18)** oluşturulur, denklemlerde w_j ağırlık matrisinin *j.*sütunu ifade etmektedir, ve denklemler boyunda w ağırlık matrisinin tüm *1.* sütunundan *H.* sütununa kadar taranır.

$$p(v) = \exp(a^T v) + \sum_{h_1\in\{0,1\}} \ldots . \sum_{h_H\in\{0,1\}} \exp\left(\sum_j h_j w_{j.}.v_s + b_j h_j\right)/Z \quad \textbf{(3.18)}$$

Her saklı işlem birimi için denklem **(3.18)**'deki toplam ifadeleri ayrı ayrı, 1. saklı işlem biriminden *H.* saklı işem birimine kadar yazılırsa açık formda denklem **(3.19)** ile ifade edilir.

$$p(v) = \exp(a^T v) +$$

$$\frac{1}{Z}\left(\sum_{h_1 \in \{0,1\}} \exp(h_1 w_{1.} . v_s + b_1 h_1)\right) \ldots \left(\sum_{h_H \in \{0,1\}} \exp(h_H w_H . v_s + b_1 h_1)\right) \quad \textbf{(3.19)}$$

Denklem **(3.19)**'da saklı işlem birimlerinin değerleri yerine yazılırsa (0 ve 1) toplam işlevinden arındırılmış **(3.20)** denklemi elde edilir.

$$p(v) = \frac{1}{Z}\exp(a^T v)\left[(1 + \exp(b_1 + w_{1.s} . v_s)) \ldots (1 + \exp(b_H + w_{H.s} . v_s))\right] \quad \textbf{(3.20)}$$

Denklem **(3.20)**'deki ifade de üstel gösterilime sahip olmayan terimler üstel ve logaritmik fonksiyon işlevlerine tabi tutularak denklem **(3.21)** elde edilir.

$$p(v) = \exp(a^T v)$$

$$\frac{1}{Z}\exp(\log(1 + \exp(b_1 + w_1 . v_s))) \ldots \exp(\log(1 + \exp(b_H + w_H . v_s))) \quad \textbf{(3.21)}$$

Denklem **(3.21)**'deki üstel (exp) işlevler denklem **(3.22)**'de tek bir üstel işlev altında toplanarak, terimler tüm saklı işlem birimleri üzerinden toplanır.

$$p(v) = \frac{1}{Z}\exp\left(a^T v + \sum_{j=1}^{H} \log\left(1 + \exp\left(b_j + w_j . v_s\right)\right)\right) \quad \textbf{(3.22)}$$

$$p(v) = \exp\left(F(x)\right)/Z \quad \textbf{(3.23)}$$

Denklem **(3.23)**'de görülen *F(x)* fonksiyonu, *x* fonksiyonunun içerdiği parametrelerin tanımladığı KBM mimarisine karşı düşen serbest enerji olarak adlandırılır.

3.3 Ön-Eğitim ile Ağ Parametrelerinin Öğreticisiz Öğrenilmesi

Konuşmacının SL-NSL durumlarının öğrenilmesi, tanımlanan üstel enerjinin en küçüklenmesi problemi olarak ele alınır ve seçilen hata fonksiyonu olan "olasılık fonksiyonunun negatif logaritmasının" en küçüklenmesi ile gerçekleştirilir. t. eğitim örneği için ortalama hata fonksiyonu denklem **(3.24)**'teki şekilde negatif logaritma kullanılarak yazılabilir. Denklem (3.24)'teki "T" eğitim öznitelik vektörlerinin toplam sayısıdır.

$$\frac{1}{T}\sum_t l\left(f\left(v^{(t)}\right)\right) = \frac{1}{T}\sum_t -logp(v^{(t)}) \quad \textbf{(3.24)}$$

Ortalama negatif log olasılık fonksiyonunu en küçükleyerek en iyi ağ parametrelerinin hesaplanmasında SGD algoritması kullanılır [30] ve denklem **(3.25)** ile görülebileceği gibi hata fonksiyonu üzerinden herhangi bir θ (ağırlıklar ya da destek terimleri) parametresinin kısmi türevi iki faz arasındaki farka bağlıdır [23].

$$\frac{\partial(-logp(v^{(t)}))}{\partial\theta} = \underbrace{E_h\left\langle\frac{\partial E(v^{(t)},h)}{\partial\theta}\middle|v^{(t)}\right\rangle}_{pozitif\ faz} - \underbrace{E_{v,h}\langle\frac{\partial E(v,h)}{\partial\theta}\rangle}_{negatif\ faz} \quad \textbf{(3.25)}$$

Pozitif faz gözleme bağlı değişirken, negatif faz sadece modele bağlıdır. Pozitif fazda saklı işlem birimleri üzerinden "beklenen kısmi türev" değeri, gözlemler koşulu altında enerjinin parametreler ile kısmı türevinin alınması ile bulunur. Negatif fazda ise, görünür işlem birimleri ve saklı işlem birimleri üzerinden modeldeki enerji fonksiyonunun kısmi türevinin beklenen değeri hesaplanır. Beklenen değer işleminin kullanılmasının nedeni saklı değerlerin bilinmemesidir.

Hesaplanmak istenen parametrenin görünür ve saklı birimler arasındaki ağırlık parametresine göre kısmi türevleri denklem **(3.26)** ile hesaplanabilir.

$$\frac{\partial E(v,h)}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}}\left(-\sum_{i\,\in\,görünür} a_i v_i \;-\; \sum_{j\,\in\,saklı} b_j h_j \;-\; \sum_{i,j} v_i h_j w_{ij}\right) \quad \textbf{(3.26)}$$

Denklem **(3.26)**'da ağırlıklardan bağımsız terimler türevi alındığında sıfırlanır ve geriye kalan terimin türevi alınırsa denklem **(3.27)** ve matris formda **(3.28)** elde edilir.

$$= \frac{\partial}{\partial w_{ij}}\sum_{i,j} v_i h_j w_{ij} = -v_i h_j \quad \textbf{(3.27)}$$

$$\nabla_w E(x,h) = -hv^T \quad \textbf{(3.28)}$$

Saklı birimler üzerinden ve koşullu olarak t. örneğin görünür işlem birimlerine bağlı beklenen değer hesaplayarak pozitif faz denklem **(3.29)** ile elde edilir. Denklem **(3.29)** tüm saklı işlem birimlerinin, *h*, değerleri (0 ve 1) üzerinden türetilmiştir ve θ değeri parametreleri (destek terimleri ve ağırlık katsayıları) gösterir. 0 değerlerinin denkleme etkisi olmadığı için işlem sonucu sadece "1" değerlerine bağlı çıkmıştır.

$$E_h\left\langle\frac{\partial(v^{(t)},h)}{\partial\theta}\middle|v^{(t)}\right\rangle = E_h\left[-v_i h_j\middle|v\right] = \sum_{h_j\in\{0,1\}}(-v_i h_j p\left(h_j\middle|v\right)) = -v_i p\left(h_j = 1\middle|v\right) \quad \textbf{(3.29)}$$

Aynı ifade denklem **(3.30)**'da ve **(3.31)**'de matris formda ifade edilmiştir. Denklem **(3.30)**, sigmoid fonksiyonu formulasyonu ile gösterimde sadeleştirilerek denklem **(3.31)** elde edilmiştir.

$$E_h[\nabla_w E(x,h) \mid v] = -h(x)x^T \quad \textbf{(3.30)}$$

$$h(x) = sigm(b + wv) \quad \textbf{(3.31)}$$

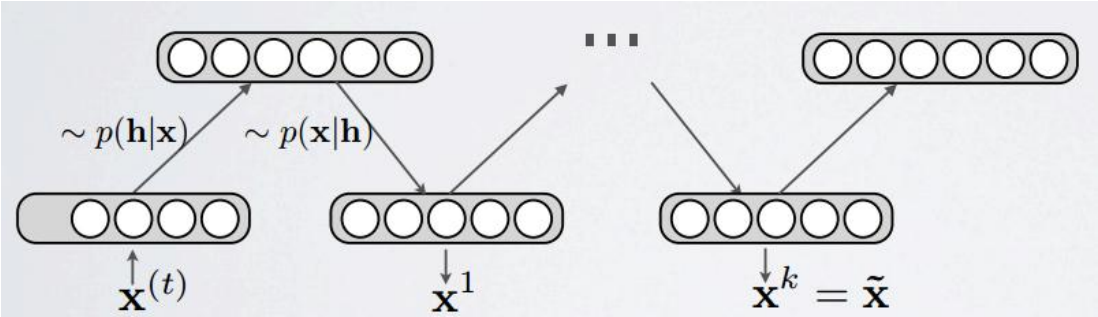
Saklı ve görünür değerler üzerinden üstel toplama işlemi gerçekleştirildiği için negatif fazın hesaplaması zordur. Bu yüzden negatif fazın belirli bir yaklaşıklıkla hesaplanabilmesi için **Geoffrey Hinton'ın** [13] karşıtlıklı ıraksay ("Contrastive Divergence", CD) algoritması kullanılır. Algoritma, 3 temel adımdan oluşur, bu adımlar:

1. Her bir örnek için döngü tekrarlanır ve döngü eğitim kümesinden ilk örnekle başlar.
2. Modellenen olasılık dağılımı üzerinden Gibbs örnekleme metodu ile «ṽ» değeri elde edilir.
3. Parametreler(ağırlık ve destek) güncellenir

olarak özetlenebilir.

İlk adımda saklı ve görünür değişkenler üzerinden çift beklenti değeri hesaplamanın işlemsel karmaşıklığı, v görünür değerinin yaklaşık değeri kullanılarak aşılır, ve saklı birimler üzerinden beklenti hesaplanır. Bu sayede denklem **(3.6)**'da paydada görülen yoğunluk fonksiyonu kestirilmiş olur.

İkinci adımda, eğitimde kullanılacak gözlem görünür işlem birimlerinin alacağı değerlerdir. Ardından tüm saklı birimler eğitimde kullanılan örneğin gözlemlenmesi koşulu altında denklem **(3.6)**'da görülen sonsal olasılık dağılımı örneklenir. Bir diğer deyişle ilk Gibbs adımında örnekleme $p(h|x = x^t)$ dağılımı üzerinden yapılır. Saklı katmandaki işlem birimleri koşullu olarak bağımsız oldukları için, bu Bernoulli rastlantı değişkenlerinin olasılıkları ($p(h_j = 1|x)$ hesaplanabilir, hesaplanan bu olasılık alınan bir tekdüze dağılımdan(0 ile 1 arasında dağılmış) çekilen bir örnekten büyük ise işlem biriminin değeri 1 olarak atanır. İkinci Gibbs adımında ise örnekleme $p(x|h)$ üzerinden yapılarak x^1 geri çatılır ve "k" Gibbs adımı sonrasında tahmini $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{v}}$ değerine ulaşılır, Şekil 3.4'te Gibbs örnekleme işlevi özetlenmiştir.



Şekil 3.4 : Gibbs örnekleme işlevi.

3.3.1 Parametrelerin güncellenmesi

Öğrenilecek bir girdi örneğe atanan olasılık değerinin arttırılması için ilgili örneğin enerji değerinin azaltılması, diğer tüm örneklerin ise enerji değerinin arttırılması gerekmektedir. Düşük enerjili örnekler bileşen (Z) fonksiyonuna en yüksek katkıyı yapan örneklerdir. Bu örneklerin enerji değerleri arttırılarak bileşen fonksiyonundaki etkileri düşürülür. Bir örneğin enerjisinin arttırılması ya da azaltılması ağırlıkların ve destek (bias) terimlerinin değerlerinin değiştirilmesi ile sağlanır. Değerlerin sürekli güncellenmesi ile doğru örneklerin olasılıkları arttırılır ve bir örneğin logaritma fonksiyonu içerisindeki türevinin değeri kolaylıkla hesaplanabilir. CD algoritmasının sonunda gözlemlenen değer (v), negatif faz için kestirim değeri ($\tilde{v}$) kullanılarak destek(bias) ve ağırlık parametreleri güncellenir. Ağırlık parametreleri denklem **(3.35)** kullanılarak güncellenmiştir.

Denklem **(3.32)**'de ağırlık matrisini güncellemek için bir ϵ öğrenme hızı ile ağırlıklandırılmış gözlemlenen eğitim örneği için kestirilen eğim değeri kullanılır.

$$W \leftarrow W - \epsilon\left(\nabla w - logp(v^t)\right) \tag{3.32}$$

Daha önce de ifade edildiği ve deklem **(3.33)**'de gösterildiği gibi eğim pozitif faz ve negatif faza bağlıdır.

$$W \leftarrow W - \epsilon\left(E_h\left[\nabla_w E\left(v^{(t)}, h\right)\middle| v^{(t)}\right] - E_{x,h}[\nabla_w E(v,h)]\right) \tag{3.33}$$

Denklem **(3.34)**'te yukarıda bahsediğidiği gibi negatif fazdaki çift beklenti değeri hesaplanmayarak yaklaşıklık kullanılmıştır.

$$W \leftarrow W - \epsilon\left(E_h\left[\nabla_w E\left(v^{(t)}, h\right)\middle| v^{(t)}\right] - E_h[\nabla_w E(\tilde{v}, h)|\ \tilde{v}]\right) \tag{3.34}$$

$$W \leftarrow W + \epsilon(h(v^t)v^{(t)^T} - h(\ \tilde{v})\ \tilde{v}^T) \tag{3.35}$$

Destek bias terimleri de denklemler **(3.43)** ve **(3.44)**'de görüldüğü gibi güncellenir.

$$b = b + \epsilon\big(h(v^t) - h(\tilde{v})\big) \quad \textbf{(3.36)}$$

$$a = a + \epsilon(v^t - \tilde{v}) \quad \textbf{(3.37)}$$

Güncellemelerden sonra ön-eğitim işlevi tamamlanır ve bir sonraki alt bölümde açıklanan eniyileme adımına geçilir.

3.4 Öğreticili Öğrenme ile Ağ Parametrelerini En İyileme

Ön-eğitim aşamasında elde edilen ağırlık katsayıları geri-yayma (backpropogation) algoritması kullanılarak sonsal (posterior) olasılıklara çevrilir ve sistemin sınıflandırıcı olarak performansını arttıran bir modele erişilir [9]. Geri-yayma algoritması, KBM kullanılarak üretici (generative) modellerle ön-eğitim uygulanmış ağ parametreleri başlangıç durumu olarak alınarak, ayırıcı (discriminative) modeller kullanarak ağırlıkların sınıflandırıcıda kullanılan en iyilenmiş değerlerine yakınsamasını sağlar. Bu tür hiyerarşik öğrenme modelleri son dönemde geliştirilen derinlikli öğrenme algoritmalarında sıklıkla kullanılmaktadır ve geri yayma ile etkinleştirilen öğrenme prosedürü sayesinde girişteki büyük veriye uyarlanmış sınıflandırıcı tasarımına olanak tanımaktadırlar [26].

Verilen eğitim setindeki her bir gözlemi, p. örnek için X_p giriş vektörü ve l_p ($l_p \in \{1, .. C\}$) sınıf belirteci olmak üzere $D_{eğitim} = \{(X_p, l_p)\}$ olarak gösterelim. Ele alınan uykulu / uykusuz ikili sınıflandırma problemi için C=2'dir. Çalışmada kullanılan giriş vektörü dokuz öznitelikten oluşan dokuz boyutlu bir vektördür. X_p sınıf belirteci ise vektörlerin uykulu ve uykusuz ikili sınıflarından hangisine dahil olduğunu belirten rakamlardır, $l_p \in \{1,2\}$.

(3.38) ve **(3.39)** denklemleri sırasıyla üretici ve ayırıcı modeller ile enbüyüklenmek istenilen ortak ve sonsal olasılık dağılımlarını göstermektedir.

$$\mathcal{L}_{\text{üretici}}\big(D_{eğitim}\big) = -\sum_{p=1}^{|D_{eğitim}|} \log\big(p\big(X_p\big)\big) \quad \textbf{(3.38)}$$

$$\mathcal{L}_{ayırıcı}\big(D_{eğitim}\big) = -\sum_{p=1}^{|D_{eğitim}|} \log\big(p\big(y_p \mid l_p\big)\big) \quad \textbf{(3.39)}$$

Gerçeklenen sistemde ilk adımda üretici bir model ile elde ettiğimiz başlangıç parametreleri, ayırıcı modelin başlangıç değerleri olarak kullanılmakta ve eğiticili

eğitim gerçekleştirilmektedir. Bu nedenle **(3.38)** denkleminin enbüyüklenmesi ile elde edilen ağırlık parametreleri **(3.39)** sonsal dağılımında başlangıç değerler olarak kullanılır. Eniyilemede kullanılan geri yayma prosedürü temel olarak istenilen çıkış vektörü ile ağın ürettiği çıkış vektörü arasındaki farkı en küçüklemeyi amaçlar, bu işlem tekrarlı olarak gerçekleşir. Ağırlıklara getirilen düzenlemeler ile girişte ya da çıkışta yer almayan ağın içerisindeki saklı işlem birimleri önemli öznitelikleri yansıtmaya başlarlar. Prosedür en alt katmandan başlayarak katman içi birimlerin durumlarını paralel olarak, üst katmanlara doğru seri olarak günceller [14].

j. işlem biriminin toplam giriş değeri x_j, y_i çıkış vektörünün doğrusal bir foksiyonudur, w_{ji} ise *i. ve j.* işlem birimleri arasındaki ağırlık olmak üzere, bu denklemler arasındaki ilişki denklem **(3.40)** ile ifade edilebilir. Kurulan ağın 3 katmanlı yapısı olduğu için *i ve j* her katmanda farklılık gösterir. Şekil (3.1)(b)'de görülen ağ için, ilk katman için $i_{max}=9$, $j_{max}=500$; ikinci katman için $i_{max}=500$, $j_{max}=500$; üçüncü katman için $i_{max}=500$, $j_{max}=2000$ değerlerini alır.

$$x_j = \sum_i y_i w_{ji} \tag{3.40}$$

İkinci aşamada gerçeklenen öğreticili eğitimde destek terimlerine 1 değeri atanarak giriş vektörünün boyutu arttırılır ve destek terimleri de algoritmaya katılır. Bu nedenle **(3.41)** denkleminde *i ve j* parametrelerinin değişimi 1. katmanda $1<i<9$, $j<j<501$ aralığındadır, diğer katmanlarda da benzer olarak çıkış parametrelerinin aralığı 1 artmış olur.

Uygulanan girişlere bağlı olarak *j. işlem biriminin* çıkış değeri , y_j, toplam giriş değerinin doğrusal olmayan bir fonksiyonu şeklinde Denklem **(3.41)**'deki gibi tanımlanır.

$$y_j = \frac{1}{1+ e^{-x_j}} \tag{3.41}$$

Geri yayma prosedürleri için denklemler **(3.40)** ve **(3.41)**'in mutlak suretle kullanılması zorunlu değildir. Türevi sınırlı değerler arasında olmak kaydıyla değişik giriş-çıkış fonksiyonları aynı ilişkiyi modelleyebilir. Fakat doğrusal olmayan bir ilişki kurmadan önce girişleri işlem birimlerine bağlayan bir doğrusal fonksiyonun varlığı öğrenme prosedürünü basitleştirmektedir.

Her iterasyonda ağa ait *j. nöronun p. giriş örneği* için ürettiği çıkış değerinin, y_{pj}'nin istenilen çıkış değeri $i_{j,p}$'ye olabildiğince yakın olması amacıyla, eniyilemede en küçüklenen hata fonksiyonu denklem **(3.42)**'deki gibi ifade edilir. Denklem **(3.42)**'de görülen $i_{j,p}$ p örneğine ait j nöronu için istenilen çıktıyı ifade etmektedir. c=1,...,C gözlenen girişin ait olabileceği sınıf etiketini göstermektedir.

$$E_p = \frac{1}{2} \sum_c \sum_j (y_{pj}(w) - i_{pj})^2 \qquad \textbf{(3.42)}$$

Denklem **(3.42)**'de verilen hata fonksiyonunu enküçükleyen ağ parametrelerinin öğrenilmesi tez kapsamında ilgili literatüre uygun olarak doğrusal olmayan bir eniyileme problemi olarak ele alınmıştır. En küçükleme işleminde farklı eğim tabanlı algoritmalar kullanılabilir, bu çalışmada hızlı ve basit olması nedeniyle eşlenik eğim (conjugate gradient) en iyileme metodu kullanılmıştır [9] [16] [15]. Eşlenik-eğim metodu "Steepest Descent" metodundan hızlı, Newton en küçükleme metoduna göre ise yavaştır. Hesian matris hesabına gerek duymadığı için sinir ağları gibi büyük matrislerle çalışılan problemlere bellek işlemleri bakımından uygundur. Çalışmada denklem **(3.42)** ile tanımlı hatanın en küçüklenmesi ile elde edilen ağırlık ve destek katsayılarının hesaplanmasının anlaşılabilmesi açısından, kullanılan eşlenik eğim algoritmasının teorik temeli aşağıdaki alt başlıklarda anlatılmaktadır.

3.4.1 Geri yayma algoritmasında eşlenik eğim metodu

Eşlenik-Eğim metodunu uygulayabilmek için bağımsız değişkenlerden, *w*, oluşan bir vektöre ve bu vektör uzayında tanımlı ve her *w* için eğimi hesaplanabilecek bir hedef fonksiyonuna, *f(w),* ihtiyaç duyulur. Geri-yayma prosedürü için bağımsız değişkenler ağırlık katsayılarıdır, *p* elemanlı eğitim kümesi üzerinden hesaplanan çıkış hatalarının normalize edilmiş toplamını hesaplayan fonksiyon ise denklem **(3.43)** ile gösterilen hedef fonksiyondur.

$$f(w) = \frac{1}{p} \sum_p E_p = \frac{1}{2p} \sum_p \sum_j (y_{pj}(w) - i_{pj})^2 \qquad \textbf{(3.43)}$$

Hedef fonksiyonun denklem **(3.44)**'de hesaplanan eğimi en iyi ağırlık katsayılarını elde etmede kullanılacaktır.

$$g(w) = \nabla f(w) = \frac{1}{p} \sum_p \nabla E_p(w) \qquad \textbf{(3.44)}$$

$$\frac{\partial E_p}{\partial w_{ji}} = \frac{\partial E_p}{\partial net_{pj}} \frac{\partial net_{pj}}{\partial w_{ji}} \quad (3.45)$$

Denklem **(3.45)**'teki $\partial E_p/\partial net_{pj}$ gösterimi j nöronunun girişindeki değişimin p örneği için ortaya çıkan hatasındaki, E_p, değişimine; $\partial net_{pj}/\partial w_{ji}$ j nöronu ve i nöronu arasındaki katsayıdaki değişimin j nöronunun girişindeki değerin değişimine, nasıl yansıdığını formüle etmektedir. net_{pj} j nöronunun girişi ve $y_{pi}(w)$, i çıkış nöronunun çıkışıdır (i giriş nöronu olsaydı $y_{pi}(w)$ giriş değeri olacaktı).

$$net_{pj} = \sum_i w_{ji}\, y_{pi}(w) \quad (3.46)$$

$$\partial net_{pj}/\partial w_{ji} = y_{pi}(w) \quad (3.47)$$

$$\partial E_p/\partial w_{ji} = \delta_{pj} y_{pi}(w) \quad (3.48)$$

Çıkış nöronları için ifade sadeleştirilirse Denklem (3.50)'deki formülasyon elde edilir. $s_j'(net_{pj})$ saklı katmandaki nöronlar için yarı-doğrusal aktivasyon fonksiyonunun türevidir.

$$\delta_{pj} = -(y_{pj}(w) - i_{pj}) s_j'(net_{pj}) \quad (3.49)$$

$$\delta_{pj} = s_j'(net_{pj}) \sum_k \delta_{pk} w_{kj} \quad (3.50)$$

Amaç fonksiyonunun eğiminin hesaplanabilmesi için eğitim kümesindeki her örneğin nöron çıkışlarını oluşturabilmesi için ağ üzerinde ileri yönlü yayılması ve daha sonra δ_{pj} değerlerinin ağ üzerinde geri yönde yayılması gerekir. Sonuç olarak da eğimin değeri, tüm örnekler üzerinden $\nabla E_p(w)$ değerinin toplanmasıyla elde edilir.

Geri yayma algoritmasında kullanılan eşlenik-eğim metodu Rasmussen'in [16] önerdiği en küçükleme kodu kullanılarak gerçeklenmiştir.

Eşlenik eğim metodunun teorik altyapısı EK.A'da detaylı anlatılmıştır.

3.5 Sınıflandırma

Şekil 3.1(b)'de gösterilen en iyileme adımının son aşamasında görüleceği gibi, ele alınan test örneği için iki sınıftan bir tanesine karar verilir.

Alt katmanlarda özyinelemeli olarak elde edilen eniyilenmiş parametrelerin bir üst katmanda kullanılması ile katmanlı bir eğitim gerçekleşir; en üst katmandaki

parametreler ve saklı işlem birimleri ile *y test örneği* için sınıf olasılıkları denklem **(3.51)**'deki gibi hesaplanabilir.

$$p(l_y = 1 \mid h;\ \theta) = softmax(\ \sum_{j=1}^{H} w_{ij} h_j + c_j\ ) \qquad \textbf{(3.51)}$$

Denklem **(3.52)**'deki $p(l \mid v)$ değerini maksimize eden eniyilenmiş parametreler ile iki sınıftan bir tanesine karar verilmiş olur.

$$p(l|v) = \sum_h e^{-E(v,l,h))} / \sum_l \sum_h e^{-E(v,l,h))} \qquad \textbf{(3.52)}$$

4. BAŞARIM TESTLERİ

Kurulan ağın eğitim ve testlerinde kullanılan ses kayıtları için *"Sleepy Language Corpus"* [7] veritabanı kullanılmıştır ve testler sonucunda elde edilen başarımlar raporlanırken aynı veritabanını kullanan çalışmalarla birlikte raporlanmıştır. Alt başlık 4.1'de eğitim, test kümeleri ve senaryoları tanıtılmış olup; alt başlık 4.2'de öğrenme ve sınıflandırma algoritmalarının farkını anlatabilmek için seçilen senaryolar üzerinden tanımadaki başarım oranları ve aynı kümeyi kullanan literatürdeki diğer çalışmalar ile karşılaştırmalı başarım oranları verilmiştir.

4.1 Uykululuk Konuşma Veri Seti ve Test Senaryoları

2011 yılında düzenlenen bir akademik yarışma olan "Speaker State Challenge" [6] [7], komite tarafından belirlenen durumlar, sesten uykululuk ve sarhoşluk durumunun belirlenmesi, için katılımcıların öznitelik-sınıflandırıcı performanslarını karşılaştırarak literatüre sunmayı amaçlamıştır. Performansların aynı referanslarla değerlendirebilmesi için de katılımcıların kullanımına testler ve eğitim için üç farklı veri kümesi sunulmuştur; eğitim veri kümesi, test veri kümesi, geliştirme veri kümesi. Üç eğitim kümesinde de farklı konuşmacılar yer alıp, testlerin olabildiğince günlük hayatı yansıtması sağlanmak istenmiştir. Eğitim veri kümesinde 20 kadın, 16 erkek toplam 36, test veri kümesinde 19 kadın 14 erkek toplam 33, geliştirme veri kümesinde ise 17 kadın 13 erkek toplam 30 konuşmacı bulunur.

Veri kümelerinde yer alan kayıtlar gırtlak özellikleri farklı, yaşları 20-52 arasında değişen ve ortalaması 24.9, standart sapması 4.2 olan 99 kişinin farklı uykululuk seviyelerindeki toplam 21 saatlik kayıtları olup, uykululuk seviyelerini belirlemek için İsveç Karolinska Enstitüsü tarafından literatüre sunulan KSS Ölçeği kullanılmıştır [10]. Ölçeğe göre uykululuk durumu 2 ana sınıf, uykululuk ve uykusuzluk, olmak üzere 10 alt sınıftan oluşmuş olup, her alt sınıf 1 ila 10 arasında bir puana sahiptir. 7.5 altı puana sahip alt sınıflar uykusuzluk sınıfını tanımlarken, 7.5 üstü puana sahip alt sınıflar ise uykululuk durumunu tanımlarlar. 10 alt sınıf puanlara göre; (1) oldukça dinç, (2) çok dinç, (3) dinç, (4) uykusu yok, (5) ne uykulu

ne uykusuz, (6) uykululuk emareleri var, (7) uykulu ama uykuyla savaşacak kadar değil, (8) uykulu ve uykuyla savaşa başladı, (9) oldukça uykuyla ciddi bir savaş içinde, (10) Ayakta kalamayacak kadar uykulu.

Konuşmaların gerçek hayata uygun senaryoların testinde kullanılabilmesi için kayıtlar arabanın içerisinde ve dersliklerde yapılmıştır. Dijital kayıtların örnekleme hızı 44.1 kHz olup 16 kHz'e düşürülmüştür, 16 bit kuantalama kullanılmış olup mikrofon ağız uzaklığı 0.3 metredir. Veritabanında bulunan kayıtlar rasgele okuma, çalışılmış okuma ve komutlar olarak üçe ayrılır. [6].

Uykululuk/Uykusuzluk probleminin araştırılması için Çizelge 4.1'de görülen 2 tip test senaryosu ve ek olarak doğrulama testleri kullanılmıştır, doğrulama testleri eğitim kümesinin hem eğitim hem testte kullanıldığı durumdur ve kullanılan kümeler, "Öbeklenmiş Eğitim Kümesi (TR1)", "Öbeklenmiş Eğitim+Geliştirme Kümesi (TR1+TR2)", "Öbeklenmemiş Test Kümesi"dir; ve kurulan ağın probleme uygunluğunu gösterebilmek için kurgulanmıştır. İlk tip senaryoda büyüklük ve konuşmacı olarak farklı eğitim kümeleri ile farklı test kümelerinin kullanıldığı durumdur, eğitimde "Öbeklenmiş Eğitim Kümesi(TR1)", "Öbeklenmiş Eğitim+Geliştirme Kümesi(TR1+TR2)", testte ise "Öbeklenmemiş Test Kümesi(TE2)" ve "Öbeklenmemiş Geliştirme Kümesi(TE1)" kullanılmıştır; kurguyla amaç kümelerdeki boyut ve kullanıcı değişimi ile sonuçlardaki tutarlılığın korunduğunun gösterilmesidir. İkinci tip senaryoda ise eğitimde "Öbeklenmemiş Test Kümesi(TE2)"nin ikiye bölünmesiyle elde edilen ilk küme "Öbeklenmemiş Test Kümesi(TR4)" ve gene "Öbeklenmemiş Test Kümesi(TE2)"nin dengeli, sınıflardan eşit oranda örnek atarak, bir şekilde sekizde birine düşürülmesiyle elde edilen "Öbeklenmemiş Test Kümesi(TR3)", testte ise "Öbeklenmemiş Geliştirme Kümesi(TE1)" ve yine "Öbeklenmemiş Test Kümesi(TE2)"nin ikiye bölünmesiyle elde edilen ikinci küme "Öbeklenmemiş Test Kümesi(TE3)" kullanılmıştır; amaç öbeklenmiş ve öbeklenmemiş kümelerdeki performans farklarını ve tutarlılığı göstermektir.

Çizelge 4.1 : Test senaryoları ve senaryolarda kullanılan kümelere dair bilgiler.

Test Senaryo	Eğitim Kümesi	Ses Kayıt Sayısı		Öznitelik Vektörü Sayısı	Kod Kelimesi Sayısı	Test Kümesi	Sınıflar		Öznitelik Vektörü Sayısı
		SL	NSL				SL	NSL	
TS-1	TR1	1241	2125	706.817	47900	TE2	851	1957	219700
	TR1+ TR2	2320	3961	1.248.710	83900	TE2	851	1957	219700
	TR1	1241	2125	706.817	47900	TE1	1079	1836	541800
TS-2	TR4	40176	36824	77000	-	TE3	57365	52235	109600
	TR3	14329	11971	26300	-	TE1	300252	241548	541800

4.2 Test Başarım Oranları

Kurulan derinlikli öğrenme ağının probleme uygunluğunun ortaya konabilmesi amacıyla 2 farklı test senaryosu ve ek olarak doğrulama testleri yapılarak her bir senaryo için birden fazla test ile başarım oranları belirtilmiştir. Başarım oranları hassasiyet (%) ve hatırlama (%) metrikleri ve bu iki metriğin ortalaması olan ağırlıklandırılmamış ortalama kullanılarak raporlanmıştır. Hassasiyet oranı, örnek üzerinden gitmek gerekirse, SL sınıfı için; SL sınıfında doğru sınıflandırılan kayıtların, sınıflandırma sonucunda "SL" olarak karar verilen tüm kayıtlara oranıdır. Hatırlama oranı ise yine SL sınıfı üzerinden gidersek sınıflandırma sonucunda doğru sınıflandırılmış kayıt sayısının, sınıflandırmaya giren tüm SL kayıtlarının sayısına oranıdır.

Doğrulama testleri eğitimde ve testte aynı kümelerin kullanıldığı testlerdir dolayısıyla eğitimde kullanılan kayıtlar, konuşmacılar testtekiler ile aynıdır ve teorik olarak en yüksek sonucun beklendiği durumdur.

Çizelge 4.2'yi 9 öznitelikten öbeklenerek oluşturulan 47900 vektörün eğitim ve test için kullanılarak elde edilen sınıflandırma sonuçları oluşturmaktadır. Test ve eğitim sonucunda %67 başarım oranı elde edilmiş olup, uykululuk ve uykusuzluk durumlarının hassasiyet ve hatırlama oranları da tutarlı bir şekilde birbirine yakın, ağırlıksız ortalamaları ise eşit çıkmıştır.

Çizelge 4.2 : Sözlük öğrenme uygulanmış eğitim kümesinde konuşmacı durum sezimi başarım oranları.

Doğrulama Testleri	TR1						
	Hata Matrisi (%)			Hatırlama (%)		Hassasiyet (%)	
	Sınıflar	SL	NSL	SL	NSL	SL	NSL
	SL	36	14	71	63	66	69
	NSL	19	31				
	Ağırlıksız Ortalama(%)	67		67		67	

Çizelge 4.3'ü ilk testtekine göre yaklaşık 2 kat büyüyerek ve 9 öznitelikten öbeklenerek oluşturulan 83900 vektörü eğitim ve test için kullanılarak elde edilen sınıflandırma sonuçları oluşturmaktadır. Test ve eğitim sonucunda yaklaşık %67 başarım oranı elde edilmiş olup, uykululuk ve uykusuzluk durumlarının hatırlama oranları arasında yaklaşık 30 puanlık fark, hassasiyet oranları arasında ise tutarlılık bulunmaktadır.

Çizelge 4.3 : Sözlük öğrenme uygulanmış genişletilmiş eğitim kümesinde konuşmacı durum sezimi başarım oranları.

Doğrulama Testleri	TR1+TR2						
	Hata Matrisi (%)			Hatırlama (%)		Hassasiyet (%)	
		SL	NSL	SL	NSL	SL	NSL
	SL	42	10	81	52	65	71
	NSL	23	25				
	Ağırlıksız Ortalama (%)	67		67		68	

İlk test senaryosu eğitimde ve testte farklı kümelerin kullanıldığı konuşmacıların kümeler arasında örtüştüğü ve örtüşmediği durumları içerir. Senaryonun amacı eğitim(öbeklenmiş) ve testteki kümelerin değişmesiyle tutarlılığın bozulmadığını, probleme önerilen çözümün gerçek hayattaki problemlere uygun olduğunu göstermektir.

Çizelge 4.4 :'ü 9 öznitelikten öbeklenerek oluşturulan 47900 vektörün eğitim ve öbeklenmeden oluşturulan 219700 vektörün test için kullanılarak elde edilen sınıflandırma sonuçları oluşturmaktadır. Eğitim sonuçları Çizelge 4.3'te raporlanmış olan eğitim kümesi ile eğitilmiştir .Test sonucunda ağırlıksız hatırlama oranı %59 çıkarken, ağırlıksız hassasiyet oranı %70 çıkmıştır. Fakat testlerde uykululuk ve uykusuzluk hatırlama oranları arasında yaklaşık 70 puanlık tutarsızlık göze çarpmaktadır.

Çizelge 4.4 : İlk test senaryosu, öbeklenmiş eğitim ve öbeklenmemiş test kümeleri başarım oranları.

	TR1<>TE2						
	Hata Matrisi (%)			Hatırlama (%)		Hassasiyet (%)	
TS-1		SL	NSL	SL	NSL	SL	NSL
	SL	50	2	96	23	58	83
	NSL	37	11				
	Ağırlıksız Ortalama (%)	61		59		70	

Çizelge 4.5'i 9 öznitelikten öbeklenerek oluşturulan 47900 vektörün eğitim ve öbeklenmeden oluşturulan 541800 vektörün test için kullanılarak elde edilen sınıflandırma sonuçları oluşturmaktadır. Eğitim sonucunda Çizelge 4.2'de verildiği üzere %67 başarım oranı elde edilmiş olup, uykululuk ve uykusuzluk durumlarının hassasiyet ve hatırlama oranları tutarlı bir şekilde birbirine yakın çıkmıştır. Test sonucunda ağırlıksız hassasiyet oranı ve hatırlama oranı %56 çıkmıştır. Önceki testlerde karşılaşılan uykululuk ve uykusuzluk hassasiyet oranları arasındaki tutarsızlık bu testte yaşanmamıştır.

Çizelge 4.5 : İlk test senaryosu, eğitimde öbeklenmiş eğitim, testte öbeklenmemiş geliştirme kümesi kullanılarak elde edilmiş başarım oranları.

	TE-1						
	Hata Matrisi (%)			Hatırlama (%)		Hassasiyet (%)	
TS-1		SL	NSL	SL	NSL	SL	NSL
	SL	33	22	60	53	61	52
	NSL	21	24				
	Ağırlıksız Ortalama (%)	57		56		56	

Çizelge 4.6'yı 9 öznitelikten öbeklenerek oluşturulan 83900 vektörün eğitim ve öbeklenmeden oluşturulan 219700 vektörün test için kullanılarak elde edilen sınıflandırma sonuçları oluşturmaktadır. Eğitim sonucunda Çizelge 4.3'te de raporlanmış olan %67 başarım oranı elde edilmiş olup, uykululuk ve uykusuzluk durumlarının hassasiyet ve hatırlama oranları tutarlı bir şekilde birbirine yakın çıkmıştır. Test sonucunda ağırlıksız hatırlama oranı %60 çıkarken, ağırlıksız hassasiyet oranı %74 çıkmıştır. Bir önceki testte olduğu gibi uykululuk ve uykusuzluk hassasiyet oranları arasında yaklaşık 70 puanlık tutarsızlık göze çarpmaktadır.

Çizelge 4.6 : İlk test senaryosu, eğitimde öbeklenmiş eğitim+geliştirme, testte öbeklenmemiş test kümesi kullanılarak elde edilmiş başarım.

	TR1+TR2 <> TE2						
	Hata Matrisi (%)			Hatırlama (%)		Hassasiyet (%)	
TS-1		SL	NSL	SL	NSL	SL	NSL
	SL	51	1	98	23	58	90
	NSL	37	11				
	Ağırlıksız Ortalama (%)	62		60		74	

İkinci test senaryosu eğitimde ve testte hem farklı hem de aynı kümelerin kullanıldığı; konuşmacıların kümeler arasında örtüştüğü ve örtüşmediği durumları içerir. Senaryo önceki senaryoların hibrit senaryosu gibi gözükse de, amacı eğitim ve testteki kümelerin öbeklenmemiş olarak kullanılmasıyla başarım oranının değişimini gözlemlemektir, böylelikle eğitim ve test kümesi seçilimi için dikkate alınacak bir sonuç ortaya konması hedeflenmektedir.

Çizelge 4.7'yi 9 öznitelikten öbeklenmeden oluşturulan 26300 vektörün eğitim ve test için kullanılarak elde edilen sınıflandırma sonuçları oluşturmaktadır. Test ve eğitim kümeleri daha önceki testlerde de kullanılan öbeklenmemiş test kümesinin uykululuk ve uykusuzluk içeren kayıtları arasındaki oran korunarak azaltılmasıyla oluşturulmuştur. Test ve eğitim sonucunda %92 başarım oranı elde edilmiş olup, uykululuk ve uykusuzluk durumlarının hassasiyet ve hatırlama oranları da tutarlı bir şekilde birbirine yakın(yaklaşık %92) çıkmıştır.

Çizelge 4.7 : İkinci test senaryosu, eğitimde ve testte aynı kümenin, azaltılmış öbeklenmemiş test kümesi, kullanılması ile elde edilen başarım.

	TR3						
	Hata Matrisi (%)			Hatırlama (%)		Hassasiyet (%)	
TS-2		SL	NSL	SL	NSL	SL	NSL
	SL	52	3	95	87	90	94
	NSL	6	40				
	Ağırlıksız Ortalama (%)	92		91		92	

Çizelge 4.8'i 9 öznitelikten öbeklenmeden oluşturulan 77000 vektörün eğitim ve öbeklenmeden oluşturulan 109600 vektörün test için kullanılarak elde edilen sınıflandırma sonuçları oluşturmaktadır. Eğitim ve test kümeleri öbeklenmemiş test

verisinin ikiye bölünmesiyle oluşturulmuştur, dolayısıyla kümeler aynı kayıtları barındırmamakla beraber ortak konuşmacıları barındırmaktadır. Eğitim sonucunda %97 başarım oranı elde edilmiş olup, uykululuk ve uykusuzluk durumlarının hassasiyet ve hatırlama oranları tutarlı bir şekilde aynı (%97) çıkmıştır. Test sonucunda ağırlıksız hassasiyet oranı ve hatırlama oranı yaklaşık %81 çıkmıştır. Uykululuk ve uykusuzluk hassasiyet oranları arasında yaklaşık olarak 15 puanlık fark ile ağırlıksız ortalamalar %82 civarında çıkmıştır ve tutarlılık göstermiştir.

Çizelge 4.8 : İkinci test senaryosu, eğitimde ve testte, öbeklenmemiş test kümesi kullanarak aynı konuşmacının farklı kayıtları testi.

<table>
<tr><td rowspan="11">TS-2</td><td colspan="7">TR4</td></tr>
<tr><td colspan="3">Hata Matrisi (%)</td><td colspan="2">Hatırlama (%)</td><td colspan="2">Hassasiyet (%)</td></tr>
<tr><td></td><td>SL</td><td>NSL</td><td>SL</td><td>NSL</td><td>SL</td><td>NSL</td></tr>
<tr><td>SL</td><td>51</td><td>2</td><td rowspan="2">97</td><td rowspan="2">97</td><td rowspan="2">97</td><td rowspan="2">97</td></tr>
<tr><td>NSL</td><td>1</td><td>46</td></tr>
<tr><td>Ağırlıksız Ortalama (%)</td><td colspan="2">97</td><td colspan="2">97</td><td colspan="2">97</td></tr>
<tr><td colspan="7">TE3</td></tr>
<tr><td colspan="3">Hata Matrisi (%)</td><td colspan="2">Hatırlama (%)</td><td colspan="2">Hassasiyet (%)</td></tr>
<tr><td></td><td>SL</td><td>NSL</td><td>SL</td><td>NSL</td><td>SL</td><td>NSL</td></tr>
<tr><td>SL</td><td>38</td><td>14</td><td rowspan="2">73</td><td rowspan="2">90</td><td rowspan="2">89</td><td rowspan="2">75</td></tr>
<tr><td>NSL</td><td>5</td><td>43</td></tr>
<tr><td>Ağırlıksız Ortalama (%)</td><td colspan="2">81</td><td colspan="2">81</td><td colspan="2">82</td></tr>
</table>

Çizelge 4.9'u 9 öznitelikten öbeklenmeden oluşturulan 26300 vektörün eğitim ve öbeklenmeden oluşturulan 541800 vektörün test için kullanılarak elde edilen sınıflandırma sonuçları oluşturmaktadır. Eğitim kümesi daha önceki testlerde de kullanılan öbeklenmemiş test kümesinin uykululuk ve uykusuzluk içeren kayıtları arasındaki oran korunarak azaltılmasıyla oluşturulmuştur. Testte kullanılan öbeklenmemiş geliştirme kümesi ile ortak kayıt yoktur ve konuşmacı bağımsızlığı sağlanmıştır. Eğitim sonucunda %92 başarım oranı elde edilmiş olup, uykululuk ve uykusuzluk durumlarının hassasiyet ve hatırlama oranları tutarlı bir şekilde aynı (ortalama %92) çıkmıştır. Test sonucunda ağırlıksız hassasiyet oranı ve hatırlama oranı %64 çıkmıştır. Uykululuk ve uykusuzluk hatırlama oranları arasında yaklaşık olarak 8 puanlık fark ile ağırlıksız ortalamalar %64 civarında çıkmıştır ve tutarlılık göstermiştir. Uykululuk ve uykusuzluk hassasiyet oranları arasında yaklaşık olarak

12 puanlık fark ile ağırlıksız ortalamalar %64 civarında çıkmıştır ve tutarlılık göstermiştir.

Çizelge 4.9 : İkinci test senaryosu, eğitimde azaltılmış öbeklenmemiş test, testte öbeklenmemiş geliştirme kümesi kullanılarak elde edilmiş başarım.

TS2	**TR3**						
	Hata Matrisi (%)			Hatırlama (%)		Hassasiyet (%)	
		SL	NSL	SL	NSL	SL	NSL
	SL	52	3	95	88	90	94
	NSL	6	40				
	Ağırlıksız Ortalama (%)	92		91		92	
	TE1						
	Hata Matrisi			Hatırlama (%)		Hassasiyet (%)	
		SL	NSL	SL	NSL	SL	NSL
	SL	33	22	60	68	70	58
	NSL	14	30				
	Ağırlıksız Ortalama (%)	63		64		64	

Çalışma sonucunda elde edilen sonuçları detaylıca karşılaştırmak için SVM ile sınıflandırma yapılmış sonuçlar kullanılacaktır. Aynı test senaryolarının literatürdeki benzer çalışmalarda [17] kullanılan SVM parametreleri ve Weka yazılım aracı ile tekrarlanması ile Çizelge 4.10'daki sonuçlar elde edilmiştir.

SVM sınıflandırıcısı ile doğrulama testlerinde ortalama %90 başarım elde edilmiş olup, derinlikli öğrenme ile elde edilen %67'lik başarımdan daha yüksektir. Aynı şekilde Senaryo-1'de SVM sınıflandırıcı ile elde edilen ortalama %73'lük başarım, derinlikli öğrenme ile elde edilen ortalama %60 başarımdan daha yüksektir. Senaryo-1'deki farklı testlerde sonuçlar SVM'de yüksek olmasına rağmen test durumları arasında alınan sonuçlar benzerlik göstererek, iki sınıflandırıcıda da (TR1+TR2)<>(TE2) testinde en yüksek sonuçlar elde edilmiştir. Senaryo-2'de iki sınıflandırıcı ortalama %80 performans göstermiş olup, TR3 <> TE1 testinde derinlikli öğrenme daha iyi sonuç vermiştir.

Çizelge 4.10 : SVM sınıflandırıcısı ile elde edilmiş test sonuçları.

	Eğitim K.	Test K.	Test Başarım Oranı						
			Hata Matrisi (%)			Hatırlama (%)		Hassasiyet (%)	
				SL	NSL	SL	NSL	SL	NSL
Doğrulama	TR1	TR1	SL	46	4	93	90	91	93
			NSL	5	45				
			Ağırlıksız Ortalama(%)	92		92		92	
	TR1 + TR2	TR1 + TR2		SL	NSL	SL	NSL	SL	NSL
			SL	48	5	91	86	88	90
			NSL	7	41				
			Ağırlıksız Ortalama(%)	89		89		89	
TS-1	TR1	TE2		SL	NSL	SL	NSL	SL	NSL
			SL	37	16	70	76	76	70
			NSL	11	36				
			Ağırlıksız Ortalama(%)	73		73		73	
	TR1 + TR2	TE2		SL	NSL	SL	NSL	SL	NSL
			SL	43	9	83	71	76	79
			NSL	14	34				
			Ağırlıksız Ortalama(%)	77		77		78	
	TR1	TE1		SL	NSL	SL	NSL	SL	NSL
			SL	33	22	60	78	77	61
			NSL	10	35				
			Ağırlıksız Ortalama(%)	68		69		69	
TS-2	TR4	TE3		SL	NSL	SL	NSL	SL	NSL
			SL	51	1	97	76	82	96
			NSL	11	36				
			Ağırlıksız Ortalama(%)	87		87		89	
	TR3	TR3		SL	NSL	SL	NSL	SL	NSL
			SL	54	0	100	99	100	100
			NSL	0	45				
			Ağırlıksız Ortalama(%)	100%		100		100	
	TR3	TE1		SL	NSL	SL	NSL	SL	NSL
			SL	19	37	33	87	76	51
			NSL	6	39				
			Ağırlıksız Ortalama(%)	57		60		64	

5. SONUÇ VE ÖNERİLER

5.1 Çalışmanın Uygulama Alanı

Uykululuk/Uykusuzluk tespiti problemi hali hazırda gerek akademik tartışmalarda gerekse de uygulama alanında kendisine yer bulan bir problemdir, özellikle dikkat gerektiren işlerin öncesinde yapılacak tespit testleri ile işi yaparken gerçekleşecek hataların azaltılması amaçlanmaktadır.

Otomotiv endüstrisi açısından çözümü önemli olan bu problem üzerine bazı firmalar hali hazırda çalışmalarını sürdürmektedir [8]. Arabaların, M2M teknolojisinin gelişmesiyle birlikte daha güçlü işlemcilere kavuşması, bu alana spesifik geliştirilecek uygulama sayısında pozitif bir katkı yapacaktır. Uygulama sayısının artmasıyla en önemli uygulamalardan biri sanal asistan uygulamaları olacak olup, ses ile iletişim kurulabilen bu asistanlar önümüzdeki dönemde sürücünün uykululuk seviyesini tespit edip koruyucu önlemleri alabiliyor olacaktır.

Sadece otomotiv sektöründe değil dikkat eksikliği dolayısıyla ciddi iş kazalarının yaşanabileceği tüm iş sahalarında uykululuk seviyesinin tespiti önem arz etmektedir. Sesten duygu analizi, insanların depresyon belirtilerini ortaya çıkarma gibi tıbbi konularda önemli bir yere sahiptir. Literatürde oldukça fazla ses tabanlı duygu analizi algoritması olmakla birlikte bu çalışmalar çoğunlukla mutluluk, sinirlilik ve benzeri kısa süreli ve ayrık duygu durum değişimlerini sezmeye yöneliktir. Oysa uykululuk sezimi orta süreli (mid-term) duygusal durum kategorisinde olup daha uzun süreli izlenmesi gereken ve gerçekte sürekli uzayda oluşan bir konuşmacı durumudur. Bu nedenle sezimi daha zordur.

Çalışma kapsamında geliştirilen derinlikli öğrenme ağı farklı problemlere uygulanabilir. Google'ın da son dönemde elindeki etiketsiz vidyoları kullanarak bilgisayarlara kedi formunu öğretme örneğinde olduğu gibi etiketsiz veri ile gösterim elde edebilen algoritmalar, kendi kendine düşünebilen ve öğrenebilen yapıların yaygınlaştırılmasında önemli bir yer tutmaktadır. İnsanın sanal ortamda bıraktığı izlerin bir hayli fazla olduğu düşünülürse, kendi kendine öğrenebilen yapıların bu

veriyi işleyip hedefi hatasız takip etmesi ve hatta bir sonraki adımını tahmin etmesi zor olmayacaktır.

Derinlikli öğrenme algoritmaları henüz çözülememiş olan beynin çalışma mekanizmasına benzer çalışma prensiplerine sahip algoritmalardır. Dolayısıyla tıpkı beynin yaptığı gibi ilk katmanda genel özellikleri öğrenirler, ikinci katman ve daha üst katmanlarda ise bir önceki katmanda öğrendikleri arasında ilinti kurarak daha detaylı bilgileri elde ederler. Aynı prensip kullanılarak, algoritmalar görüntü işlemeden, ses işlemeye kadar eğitim ve sınıflandırma basamaklarına sahip tüm uygulamalarda kullanılabilir.

5.2 Tez Sonuçları ve Tartışma

Tez kapsamında yapılan testler ile literatürde bu problemin araştırılması için kullanılan bir veritabanından elde edilen [7] ses kayıtlarından çıkarılan özniteliklerin eğitilmesi ve sınıflandırılması sonucu kurulan ağın probleme uygunluğu sınanmıştır. Problemin özel bir problem olması dolayısıyla yaygın kullanılan tek veritabanı olan "SLC Corpus" Almanca konuşma kayıtlarını içermektedir.

Kurulan eğitim-sınıflandırıcı yapısının ilk defa bu tarz bir problemde kullanılacak olması dolayısıyla, başarım testleri de literatürde kabul gören test senaryolarının yanında raporlanmıştır. "Train vs Train" performansına dayalı doğrulama testleri sonucunda ortalama %68 başarım oranı elde edilmiştir ve iki sınıf için elde edilen hatırlama oranları SL için %66, NSL için %70 olup tutarlı sonuçlar vermektedir.

Doğrulama testlerinin sonucunda elde edilen olumlu sonuçlardan sonra literatürdeki diğer çalışmaları da referans alarak 3 başarım kıstası üzerinde durulmuştur, sonuçlar:

1. "Train vs Test" testlerinde,
    a. Farklı konuşmacılar ve
    b. Farklı konuşmacıların farklı kayıtları testlerinde de başarım göstermeli
2. Eğitim kümesinin büyüklüğünün radikal değişimlerinde sonuçlar tutarlılık göstermeli
3. SL ve NSL hatırlama oranları arasında toplam sınıflandırma sonuçları yüksek olsa bile uçurum olmamalıdır.

Bu kıstaslar referans alınan çalışmalarda [17] da kullanıldığı şekilde öbeklenmiş ve öbeklenmemiş [7] eğitim kümeleri ile yapılan eğitimlerin ardından test kümesi üzerinde elde edilen sınıflandırma sonuçlarında incelenmiştir. Referans alınan çalışmada [17] orijinal öznitelik vektörleri yerine eğitim kümesinin öbeklenmesi ile elde edilen kod vektörleriyle yapılan eğitimin öbekleme sonucu arttırıcı yönde etki yapmasına rağmen, derinlikli öğrenmenin öbeklemeye detay bilgiyi kaybettirdiğinden olumlu tepki vermeyeceği düşünülmüş olup, test senaryolarına öbeklenmemiş verideki sonuçlar da eklenmiştir.

Farklı konuşmacıların kullanıldığı birinci test senaryosunda (TS-1) hatırlama oranlarında yaklaşık %70'lik başarım oranı elde edilmiştir, bu haliyle SVM gibi "Train vs Train" performansları yüksek sınıflandırıcılarda "Train vs Test"e geçişte yaşanan dramatik azalışlar ile karşılaşılmamıştır.

Test senaryosu 1 kapsamında, eğitim kümesi büyütüldüğünde başarım %74'e çıkmıştır, ek olarak SL ve NSL sınıf bağımlı hatırlama oranları arasındaki fark bir önceki test ile karşılaştırınca tutarlı bir şekilde korunmuştur.

Çalışma kapsamında en düşük sonuç, literatürde de en zor küme olarak raporlanan kümenin test kümesi olarak ele alınmasıyla elde edilmiş olup ağırlıksız hatırlama oranı %56 çıkmıştır (Çizelge 4.5) ve SL, NSL sınıfları içerisindeki hassasiyet ve hatırlama oranları yine tutarlılık gösterip 10 puanlık banda oturmuşlardır. Bu test kümesi için sınıflandırma başarımı Çizelge 4.10 ile raporlanan SVM sonuçları arasında da en düşük çıkan başarım olmuştur.

En düşük başarımın alındığı test kümesi kullanılarak, öbeklenmemiş test kümesinde bu sefer yine aynı şekilde farklı konuşmacıları içeren başka bir küme ile eğitim gerçekleştirildiğinde (Çizelge 4.9) sonuçlar biraz daha yükselerek %64'e çıkmış, SL ve NSL sınıfları içerisindeki başarım oranları arasındaki fark 10 puan bandına oturmuştur.

Aynı konuşmacıların farklı kayıtları kullanıldığında hassasiyet ve hatırlama oranları ortalama %82'ye çıkıp; SL ve NSL sınıfları içerisindeki başarım oranları arasındaki tutarlılık korunmuştur.

Öbeklenmemiş veri kümesi ile "Train vs Train" sınıflandırma sonuçlarında hassasiyet ve hatırlama oranları ortalama %92 olarak raporlanmıştır (Çizelge 4.7),

SL ve NSL sınıfları içerisindeki hatırlama başarım oranları arasındaki fark 5 puan bandına oturmuştur.

Elde edilen sonuçlar ışığında ortaya çıkan sonuç, öbeklenmemiş kümeler ile yapılan testler sonucunda daha yüksek konuşmacı uykululuk durum sezimi başarımı elde edildiğidir.

Çizelge 4.10'da verilen SVM ile elde edilen başarımın doğrulama testlerinde ve senaryo 1'de bitirme çalışmasında derinlikli öğrenme ile elde edilen sınıflandırma başarımından daha iyi olduğu görülmektedir. Derinlikli öğrenme ile elde edilen başarımın daha iyi yapılar (katman sayısı, işlem birimi sayısı vb.) kurularak ve daha iyi donanımlar kullanılarak iyileştirilebilmesinin olanaklı olabileceği düşünülmektedir. Öte yandan, derinlikli öğrenme metodunun eş zamanlı öğrenmeye daha uygun olma avantajı vardır.

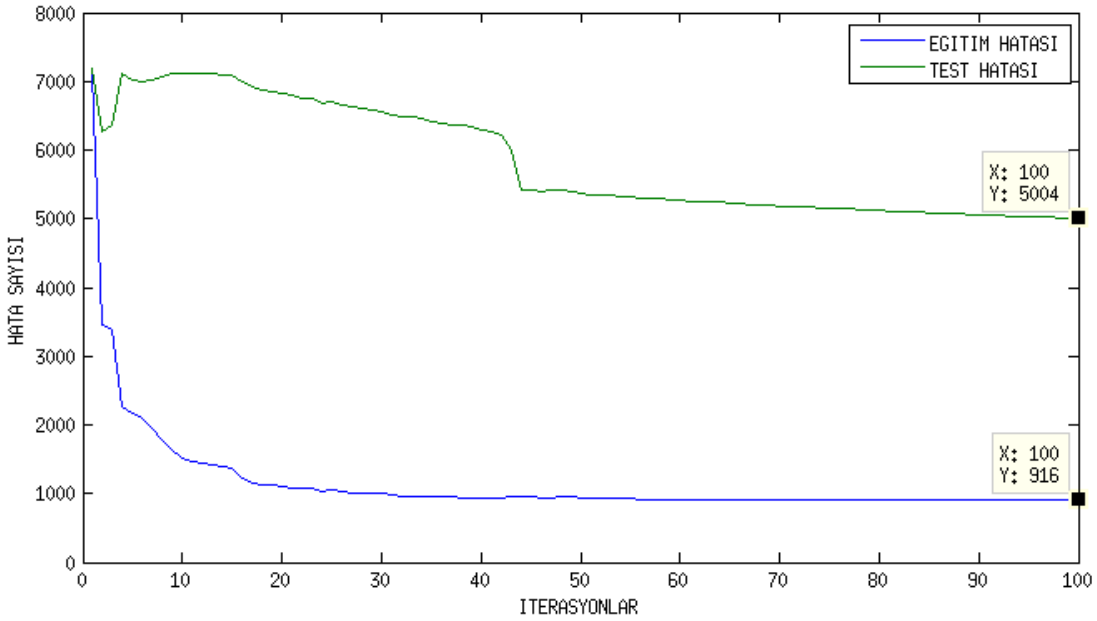
Literatürdeki diğer çalışmalar ile karşılaştırma yapabilmek için alınan sonuçlar tabloya döküldüğünde Çizelge 5.1 ortaya çıkmaktadır. Literatürde verilen sonuçlar cümle bazlı başarımı raporlarken, tez çalışması kapsamında alınan sonuçlar 43 ms'lik çerçeveler bazında hassasiyet başarımını raporlamaktadır. İlk test senaryosu ile elde edilen başarım oranları sonucunda literatürdeki sonuçların altında kalınırken, ikinci test senaryosu sonucunda literatürdeki en iyi performansa yakın bir başarım oranı elde edilmiştir.

Çizelge 5.1 : Literatür ve tez sonuçlarının karşılaştırılması.

Çalışmalar	Acc (%)					
	TRAIN vs DEV			TRAIN+DEV vs TEST		
	SL	NSL	UA	SL	NSL	UA
Tez Çalışması (Derinlikli Öğrenme)	60	53	56	58	90	74
ITU MSPPR Lab. (SVM) [17]	89.1	97.2	93.2	79.9	80.1	80.0
Interspeech 2011 Kazanan (Hibrid) [28]	60.3	75.7	68.0	64.2	79.1	71.7
Interspeech 2011 Taban Seviye (SVM) [7]	NA	NA	67.3	NA	NA	70.3

Yapay sinir ağları temelli derinlikli öğrenme algoritmaları parametrelere ve iterasyon sayılarına oldukça duyarlı algoritmalardır. Öte yandan kullanılan derinlikli öğrenme standart bir yapay sinir ağı yapısı olmayıp yönlü olmayan graf tabanlı yöntemlerle bütünleşmiştir. Daha iyi bellek kullanımı olan daha güçlü işlem birimlerinde daha yüksek sonuçlar elde edilebileceği umulmaktadır. Çalışma kapsamında elde edilen sonuçlar teorik üst limit kısıtı altında değil, işlemsel kısıtlar altında elde edilen sonuçlardır. Şekil 5-1'de görülen hata oranları senaryo-2 için elde edilen hata

oranları olup, senaryo-2'de aynı konuşmacının farklı kayıtları eğitimde ve testte (TR4 <> TE3) kullanılarak Çizelge 4.8'deki sonuçlar elde edilmişti. Parametrelerin eniyilenmesi sırasında yapılan 100 iterasyonda eğitim ve test kümelerinde yapılan hatalar çizdirildiğinde Şekil 5.1'deki grafik elde edilmektedir. Eğitimde daha yüksek başarımın (%97) elde edildiği bu senaryoda, 20 iterasyon sonunda eğitim kümesi minimum hata değerine ulaşırken, test kümesi 100 iterasyon sonunda dahi hata sayısını her iterasyonda azaltabilir durumdadır. Kurulan yapı hata sayıları üzerinden bir değerlendirmeye tabi tutulduğu takdirde, eğitim kümesi üzerinden gerçekleştirilen eniyilemenin testte olumlu sonuçlar doğurduğu ve hata sayıları arasında pozitif korelasyon olduğu görülmektedir.



Şekil 5.1 : Eğitim ve test sırasında alınan hata sayılarının karşılaştırılması.

Donanım teknolojisinde yaşanan gelişmeler ile derinlikli öğrenme algoritmalarının gerçek zamanlı öğrenmeye her geçen gün daha elverişli hale geleceği varsayımı altında, kendi kendine karar verebilen yapılar kurulabilmesi için bu algoritmalar değerli algoritmalardır. Daha güçlü donanımlarda test senaryoları denenerek teorik üst limite ulaşıldığında bu gerçek daha iyi anlaşılacaktır.

KAYNAKLAR

[1] R. Cowie et al., "Emotion Recognition in Human-Computer Interaction," *IEEE Signal Processing Magazine*, vol. 18, pp. 32-80, January 2011.

[2] J. Krajewski and B. Kröger, "Using Prosodic and Spectral Characteristics for Sleepiness Detection," in *INTERSPEECH 2007*, Antwerp, Belgium, 2007, pp. 1841-1844.

[3] F. Schiel and C. Heinrich, "Laying the Foundation for In-Car Alcohol Detection by Speech," in *INTERSPEECH 2009*, Brighton, UK, 2009, pp. 983-986.

[4] Mehmet Cenk Sezgin, Bilge Gunsel, and Gunes Karabulut Kurt, "Perceptual Audio Features for Emotion Detection," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2012:16, May 2012.

[5] F. Weninger, J. Krajewski, A. Batliner, and B. Schuller, "The voice of leadership: models and performances of automatic analysis in online speeches," *Affective Computing, IEEE Transactions*, vol. 3, no. 4, pp. 496-508, June 2012.

[6] B. Schuller et al., "Medium-Term Speaker States - A Review on Intoxication, Sleepiness and the First Challenge," in *Computer Speech and Language, Special Issue on Broadening the View on Speaker Analysis*, 2013, pp. 48-55.

[7] Björn Schuller, Anton Batliner, Stefan Steidl, Florian Schiel, and Jarek Krajewski, "The INTERSPEECH 2011 Speaker State Challenge," in *Proc. INTERSPEECH 2011*, Florence, Italy, 2011.

[8] Llc Ford Global Technologies, "System and method for establishing acoustic metrics to detect driver impairment," WO2013023032 A1, February 14, 2013.

[9] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science Magazine*, vol. 313, no. 5786, pp. 504-507, July 2006.

[10] T. Akerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *International Journal of Neuroscience*, no. 52, pp. 29-37, 1990.

[11] International Telecommunications Union, "Method for objective measurements of perceived audio quality," International Telecommunications Union, Standard BS.1387-1, 2000.

[12] Geoffrey Hinton, "A Practical Guide to Training Restricted Boltzmann Machines," University of Toronto, Toronto, Guide UTML TR 2010-003, 2010.

[13] G. E. Hinton, "Training Products of Experts by Minimizing Contrastive Divergence," *Neural Computation*, no. 14, pp. 1771-1800, 2002.

[14] D.E. Rumelhart, G. E Hinton, and R. J. Williams, "Learning Representations by Back-Propagating Errors," *Nature*, vol. 323, pp. 533-536, October 1986.

[15] E. M. Johansson, F. U. Dowla, and D. M. Goodman, "Backpropagation Learning For Multilayer Feed-Forward Neural Networks Using the Conjugate

Gradient Method," *International Journal of Neural Systems*, vol. 2, no. 4, pp. 291-301, January 1992.

[16] Carl Edward Rasmussen. (2006, April) Minimize Function. [Online]. http://learning.eng.cam.ac.uk/carl/code/minimize/minimize.m

[17] Gunsel B, Sezgin C, and Krajewski J, "Sleepiness detection from speech by perceptual features," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference*, Vancouver, BC, 2013, pp. 788-792.

[18] E. M. Schmidt and Y.E. Kim, "Learning emotion-based acoustic features with deep belief networks," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2001 IEEE Workshop*, New Paltz, NY, 2011, pp. 65-68.

[19] A. R. Mohamed, G.E Dahl, and G.E. Hinton, "Deep belief networks for phone recognition," in *NIPS 22 workshop on deep learning for speech recognition*, 2009.

[20] L Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, February 1989.

[21] Alex (Sandy) Pentland, *Honest Signals*. US: The MIT Press, 2010.

[22] F Eyben, M. Wollmer, and B Schuller, "OpenEAR — Introducing the munich open-source emotion and affect recognition toolkit ," in *Affective Computing and Intelligent Interaction and Workshops*, Genea, Switzerland, 2009, pp. 1-6.

[23] Hugo Larochelle and Yoshua Begio, "Classification using discriminative restricted boltzmann machines," in *ICML '08 Proceedings of the 25th international conference on Machine learning*, New York, USA, 2008, pp. 536-543.

[24] Ruslan Salakhutdinov and Geoffrey E. Hinton, "Deep boltzmann machines," in *International Conference on Artificial Intelligence and Statistics*, Las Vegas, USA, 2009, pp. 448-455.

[25] Martin Moller, "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning," *Neural Networks*, vol. 6, pp. 525-533, 1993.

[26] G.E Hinton, "Learning multiple layers of representation," *Trends in Congnitive Sciences*, vol. 11, no. 10, pp. 428-434, 2007.

[27] A. Stuhlsatz, C. Meyer, F. Eyben, T. ZieIke, and B. Schuller, "Deep neural networks for acoustic emotion recognition: Raising the benchmarks," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, Prague, 2011, pp. 5688 - 5691.

[28] Dong-Yan Huanga, Zhengchen Zhangb, and Shuzhi Sam Geb, "Speaker state classification based on fusion of asymmetric simple partial least squares (SIMPLS) and support vector machines," *Computer Speech & Language*, vol. 28, no. 2, pp. 392-419, March 2014.

[29] Y LeCun, D Touretzky, G Hinton, and T Sejnowski, "A theoretical framework for Back-Propagation," in *Proceedings of the 1988 Connectionist Models Summer School*, Pittsburgh, 1988, pp. 21-28.

[30] Yoshua Bengio and Olivier Delalleau, "Justifying and Generalizing Contrastive Divergence," Universite de Montreal, Montreal, Technical Report 2007.

[31] J. Krajewski, "The Center of Interdisciplinary Speech Science," University of Wuppertal, Germany,.

EKLER

EK A: Eşlenik Eğim Algoritması

EK A

Eşlenik Eğim Algoritması:

Eşlenik Yönler: Eşlenik eğim algoritmasının anlaşılabilmesi için bu bölümde [15] çalışmasından yararlanarak eşlenik-eğim teorisi üzerinde durulacaktır.

En küçüklenmek istenen amaç fonksiyonunun ikinci dereceden bir fonksiyon olduğu kabul edilirse, *f(x)* fonksiyonunun minimum değeri sıfır eğim noktasındadır, $\nabla f(x) = Qx - b = 0$. En küçük nokta, *x**, doğrusal fonksiyonun çözümü olmuş olur, Qx*=b.

$$f(x) = \frac{1}{2} x_n^T Q_{nxn} x_n - b^T x_n \tag{A.53}$$

Aynı şekilde 2. dereceden fonksiyonun çözümünü bulabilmek için n tane sıfırdan farklı değerler alan n tane Q-konjuge vektöre ihtiyaç duyarız. Aşağıdaki özelliği sağlayan vektörler Q-konjuge vektör olarak tanımlanır. Gösterilebilir ki Q-konjuge vektörler doğrusal bağımsızdırlar ve x vektör uzayını gererler.

$$d_i^T Q d_j = 0, i \neq j \tag{A.54}$$

Verilen Q-konjuge vektörleri kullanılarak x_0 noktasından *x** noktasına varılmasını temsil eden vektör Q-konjuge vektörlerin ağırlıklı, α_i, kombinasyonu ile oluşturulabilir.

$$x^* - x_0 = \sum_{i=0}^{n-1} \alpha_i d_i \tag{A.55}$$

Skaler ağırlık katsayılarını bulabilmek için denklem $d_j^T Q$ ile çarpılır, ve Qx*=b dönüşünümü yapılır.

$$d_j^T [b - Q x_0] = \sum_{i=0}^{n-1} \alpha_i d_j^T Q d_j \tag{A.56}$$

Q-konjuge vektörlerin sağladığı avantaj ile n bilinmeyen için n doğrusal denklem çözülmeden, skaler ağırlık katsayıları hesaplanır.

$$\alpha_j = \frac{d_j^T [b - Q x_0]}{d_j^T Q d_j} \tag{A.57}$$

Denklem **(A.5)** kullanılarak hedef, *x*, için özyinelemeli bir ifade yazılabilir.

$$x_k = x_0 + \sum_{j=0}^{k-1} \alpha_j d_j \tag{A.58}$$

$$x_{k+1} = x_k + \alpha_k d_k \tag{A.59}$$

$x_n = x^*$olmak üzere x_0 noktasından x^* hedefine hareket n farklı yöne doğru hareket serileri olarak ifade edilebilir. Denklem **(A.3)** ve **(A.6)** kullanılarak $d_k^T Q x_k = d_k^T Q x_0$ eşitliği gösterilebilir. $g_k = \nabla f(x_k) = Q x_k - b$ ifadesi ve **(A.5)** denklemi kullanılarak ağırlık ifadesinin güncel gösterimi elde edilir.

$$\alpha_k = \frac{d_k^T g_k}{d_k^T Q d_k} \tag{A.60}$$

(A.7) ve **(A.8)** denklemleri eşlenik yönler algoritmasını oluşturur ve algoritma kullanılarak pozitif tanımlı ikinci dereceden fonksiyon n adımda çözülebilir. Algoritmanın önemli bir özelliği güncel eğim değerinin önceki yön vektörlerine ortogonal olmasıdır.

$$g_{k+1} - g_k = Q[x_{k+1} - x_k] = \alpha_k Q d_k \tag{A.61}$$

(EK.A.3,8,9) denklemleri ve d_k iç çarpımı kullanılarak ortogonallik ispatlanır.

$$g_{k+1}^T d_k = d_k^T g_k + \alpha_k d_k^T Q d_k = 0 \tag{A.62}$$

$$[g_{k+1} - g_k]^T d_j = 0, \;\; j = 1,2, \dots\dots, k-1 \tag{A.63}$$

Ortogonallik özelliğinin önemi her adımda **(A.6,7,8)** denklemleriyle elde edilen x_{k+1} değerinin yön vektörleri tarafından gerilen alt uzayda amaç fonksiyonunu, *f(x)*, en küçüklediğini gösterir ve $g_{k+1}^T d_k = 0$ olduğu için, **(A.12)** denklemindeki en küçükleme fonksiyonunun çözümü α_k olur.

$$\min_\alpha f(x_k + \alpha d_k) \tag{A.64}$$

$f(x_k + \alpha d_k)$ fonksiyonunun minimumu $\partial f(x_k + \alpha d_k)/\partial \alpha = 0$ türevini sağlayan noktadadır.

$$\frac{\partial f(x_k + \alpha d_k)}{\partial \alpha} = \nabla f(x_k + \alpha d_k)^T d_k = g^T d_k \tag{A.65}$$

Eşlenik-Eğim Algoritmasının Adımlar

Önceki bölümde anlatılan yön vektörlerini üreten çok sayıda eşlenik yön algoritması vardır. Çalışmada kullanılan algoritma ise **Eşlenik Eğim** algoritmasıdır. Başlangıç yön vektörü olarak başlangıç noktasındaki negatif eğim seçilir, $d_0 = -g_0$, diğer yön vektörleri elde edilen eğim ve yön vektörünün doğrusal kombinasyonları ile elde edilir.

$$d_{k+1} = -g_{k+1} + \beta_k d_k \quad \textbf{(A.66)}$$

$$d_{k+1}^T Q d_k = [-g_{k+1} + \beta_k d_k]^T Q d_k = 0 \quad \textbf{(A.67)}$$

$$\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k} \quad \textbf{(A.68)}$$

(A.14, A.16) denklemleri kullanılarak elde edilen yön vektörlerinin, Q-konjuge setini oluşturdukları tümevarım yöntemiyle ispatlanabilir. $d_1^T Q d_0 = 0$ denklemini sağlayacak bir β_0 seçilir, *k<n-2* olmak üzere tüm $j < i \leq k$ değerleri için $d_i^T Q d_j = 0$ olduğunu kabul edelim. Kabulun doğru olabilimesi için tüm $j < i \leq k+1$ değerleri için $d_i^T Q d_j = 0$ olmalıdır. Kabul altında ilk olarak ilk *k+1* eğimin ortogonal kümeyi oluşturduğu ispatlanacaktır. Denklem **(A.14)** kullanılarak d_p 'nin ilk *p+1* eğimin doğrusal kombinasyonu olduğu gösterilebilir.

$$d_p = -g_p + \sum_{q=0}^{p-1} \gamma_{p,q} g_q \quad \textbf{(A.69)}$$

İlk *k* yön vektörü Q-konjuge varsayıldığı için Denklem **(A.11)** ve g_m ile çarpılarak **(A.17)** denklemi kullanılarak **(A.18)** denklemi elde edilir, $d_0 = -g_0$ olduğundan, $p < m \leq k+1$ için ${g_m}^T g_0 = 0$ olur.

$${g_m}^T g_p = \sum_{q=0}^{p-1} \gamma_{p,q} {g_m}^T g_q \quad , p < m \leq k+1 \quad \textbf{(A.70)}$$

Tanıt, **(A.14)** denklemi $Q d_j$ ile genişletilip $j < k$ için $d_{k+1}^T Q d_j = 0$ olduğu gösterilerek tamamlanır.

$$d_{k+1}^T Q d_j = -{g_{k+1}}^T Q d_j + \beta_k {d_k}^T Q d_j \quad \textbf{(A.71)}$$

Kabulden dolayı eşitliğin sağ tarafındaki 2. Terim sıfıra eşittir, ilk terim ise **(A.9)** denkleminden, $-{g_{k+1}}^T [g_{j+1} - g_j]/\alpha_k$, eğimler ortogonal olduğu için sıfırdır. Tanıt gerçekleştirildikten sonra eşlenik eğim algoritması Çizelge A.1'deki şekilde özetlenebilir.

Çizelge A.1 : Eşlenik eğim algoritmasının özeti.

1.k=0 olarak ata ve başlangıç noktası olan x_0 değerini belirle
2. $g_0 = \nabla f(x_0)$.
Eğer $g_0 = 0$ ise dur,
Aksi takdirde $d_0 = -g_0$ atamasını yap
3. $\alpha_k = -\frac{d_k^T g_k}{d_k^T Q d_k}$
4. $x_{k+1} = x_k + \alpha_k d_k$

Çizelge A.1 (devam) : Eşlenik eğim algoritmasının özeti.

5. $g_{k+1} = \nabla f(x_{k+1})$. Eğer $g_{k+1} = 0$ ise dur 6. $\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k}$ 7. $d_{k+1} = -g_{k+1} + \beta_k d_k$ 8.Eğer k=n-1 ise dur, Aksi takdirde $k \leftarrow k + 1$ atamasını yap ve 3. adıma dön

Eşlenik eğim algoritması yukarıda gösterildiği gibi pozitif tanımlı ikinci dereceden bir fonksiyonun *n* adımda en küçüklenmesi amacıyla kullanılır. **(A.1)** denklemi ikinci dereceden Taylor serisi açılımı olarak yorumlanırsa algoritma doğrusal olmayan fonksiyonaları da kapsayacak şekilde geliştirilebilir. Fakat bu tip problemlerde **Q** hesaplanması işlem karmaşıklığı açısından büyük bir maliyet oluşturmaktadır. Algoritmanın verimini arttırmak için α_k ve β_k hesaplamaları Q matrisi kullanılmadan gerçekleştirilicektir. Bu yüzden **(A.8)** denklemindeki α_k kapalı form gösterimi yerine "doğru arama prosedürü", denklem **(A.16)**'daki β_k hesaplaması için de Polak-Ribiere gösterimi kullanılacaktır. Böylelikle Q matrisi hesaplanmadan iterasyonlarda sadece fonksyion ve eğim değerleri kullanılarak verimli bir eşlenik-eğim algoritması ortaya konmuş olur. Çizelge A.1'deki 5. Adım gerçekleştirilmesi zor bir adım olduğu için algoritmada Wolfe-Powell, Denklem **(A.20)**, durdurma kriteri kullanılmıştır.

$$\beta_k = \frac{g_{k+1}^T [g_{k+1} - g_k]}{g_k^T g_k} \qquad \textbf{(A.72)}$$

<u>Doğru Arama</u>

"Doğru arama"nın amacı $f(x + \alpha d)$ fonksiyonun α'ya göre en küçüklemektir. α değişirken *x ve d* sabit tutulmaktadır, böylelikle *x*'in *n* boyutlu vektör uzayında $x + \alpha d$ bir doğru oluşturur ve "doğru arama" ismi buradan gelir. Doğru arama minimum değerinin bulunacağı aralığı sınırlamayı, ve sonrasında eğim bilgisini kullanarak ikinci ya da üçüncü dereceden eğri oturtularak minimum noktası kestirilir, ve iterasyonlar belirli bir yaklaşıklığa varınca durdurulur ve performansta belirleyici bir rol üstlenir.

ÖZGEÇMİŞ

Ad Soyad: Canberk Hacıoğlu

Doğum Yeri ve Tarihi: Edirne-05.05.1988

Adres: ITU Ayazağa Kampüsü, ARI-3 Teknokent Binası Kat-2
Vodafone Teknoloji Hizmetleri

E-Posta: canberkhacioglu@gmail.com

Lisans: İTÜ Telekomünikasyon Mühendisliği

İTÜ Bilgisayar Mühendisliği

Mesleki Deneyim ve Ödüller:

(2011- 2013) IT Planlama Uzmanı

(2013-2014) IT Tedarikçi Yönetimi Kıdemli Uzmanı

Yayın Listesi:

Sezgin, C., Gunsel, B., & Hacioglu, C. (2012, April). Audio emotion recognition by perceptual features. In Signal Processing and Communications Applications Conference (SIU), 2012 20th (pp. 1-4). IEEE.