

**SAKLI MARKOV MODELLERİNDE ÜÇ TEMEL
PROBLEMİN İNCELENMESİ**

**INVESTIGATION OF THREE BASIC PROBLEMS IN
HIDDEN MARKOV MODELS**

ZÜLEYHA SEMERCİ

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

İstatistik Anabilim Dalı İçin Öngördüğü

YÜKSEK LİSANS TEZİ

olarak hazırlanmıştır.

2006

Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Bu çalışma jürimiz tarafından **İSTATİSTİK ANABİLİM DALI** 'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan :.....
Prof. Dr. Ahmet YALNIZ

Üye(Danışman) :.....
Doç. Dr. Gül ERGÜN

Üye :.....
Yrd. Doç. Dr. İbrahim ZOR

ONAY

Bu tez, /..... / tarihinde Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından kabul edilmiştir.

.... /..... /.....

Prof. Dr. Ahmet R. ÖZDURAL
FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRÜ

Canım ANNEM ve BABAMA,
bana verdikleri her Őey iin...

SAKLI MARKOV MODELLERİNDE ÜÇ TEMEL PROBLEMİN İNCELENMESİ

Züleyha SEMERCI

ÖZ

Saklı Markov modelleri özellikle durağan olmayan konuşma sinyal süreçlerinde kullanılan istatistiksel modellerdir. Gerçek sinyallerin, sinyal modelleriyle tanımlanması önemli bir araçtır. İyi bir sinyal modeli, sinyal kaynağı elimizde olmaksızın, kaynakla ilgili pek çok şeyi öğrenmemizi sağlar. İstatistiksel modellemelerde sinyaller, stokastik süreç gibi düşünülerek oldukça iyi karakterize edilebilir ve stokastik sürecin parametreleri yine başarılı bir şekilde kestirilebilir.

Tez çalışmasında, saklı Markov modellerinin genel yapısı incelenmiş, model çeşitleri hakkında genel bir bilgi verilmiştir. Saklı Markov modellerinin tam olarak anlaşılabilmesi ve modellerin uygulamalarda kullanışlı bir hale getirilmesi için çözülmesi gereken üç temel problem ele alınmıştır.

Anahtar Kelimeler: Markov Zincirleri, Saklı Markov Modelleri

Danışman: Doç. Dr. Gül ERGÜN, Hacettepe Üniversitesi, İstatistik Bölümü,
İstatistik Anabilim Dalı

INVESTIGATION OF THREE BASIC PROBLEMS IN HIDDEN MARKOV MODELS

Züleyha SEMERCI

ABSTRACT

Hidden Markov model is a statistical model especially used in nonstationary speed processes and is an important tool for defining the true signals by a signal model. An appropriate signal model allows us to get more information about the source of the signals when the source is not clear. The signals can be characterized as a stochastic process and the parameters of the process can be successfully estimated in statistical modelling.

In this study, the general structure of hidden Markov models are examined and some information are given for the types of the models. The three important problems are investigated here for understanding the model's structure clearly.

Key Words: Markov Chain, Hidden Markov Models

Advisor: Assoc. Prof. Dr. Gül ERGÜN, Hacettepe University, Department of Statistics.

TEŐEKKÜR

Bu alıőmanın her aőamasında deęerli katkıları ve yapıcı eleőtirileriyle yön veren, desteęini hiçbir zaman esirgemeyen deęerli danıőmanım Do. Dr. Gül ERGÜN' e gösterdięi sabır ve teővikleri için sonsuz teőekkür ederim. İyi ki varsınız.

Desteklerimi hep yanımda hissettięim sevgili AİLEME teőekkür ederim.

alıőma boyunca gösterdikleri anlayıő ve teővikleri için dostlarıma teőekkür ederim.

İÇİNDEKİLER DİZİNİ

Sayfa

ÖZ.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
İÇİNDEKİLER DİZİNİ.....	iv
ŞEKİLLER DİZİNİ.....	v
1. GİRİŞ.....	1
2. OLASILIKSAL SÜREÇLER.....	3
2.1. Giriş.....	3
2.2. Olasılıksal Sürecin Tanımı.....	3
2.2.1. Bağımsız artmalı süreç.....	4
2.2.2. Durağan süreçler.....	5
2.3. Markov Sürecinin Tanımı.....	5
3. KESİKLİ PARAMETRELİ MARKOV ZİNCİRLERİ.....	6
3.1. Geçiş Olasılıkları.....	6
3.2. Olasılık vektörü ve Olasılık Matrisi.....	7
3.3. Geçiş Matrisi.....	9
3.4. Kesikli Markov Süreci	10
4. SAKLI MARKOV MODELLERİ.....	13
4.1. Saklı Markov Modelini Oluşturan Unsurlar.....	18
4.2. Saklı Markov Modelinin Varsayımları.....	21
5. SAKLI MARKOV MODELLERİNİN ÜÇ TEMEL PROBLEMİ.....	22
5.1. Saklı Markov Modellerinin Üç Ana Probleminin Çözümü.....	23
5.1.1. 1. Problemin çözümü.....	23
5.1.2. 2. Problemin Çözümü.....	31
5.1.3. 3. Problemin Çözümü.....	34
6. SONUÇLAR.....	39
KAYNAKLAR.....	40
EK.....	41

ŞEKİLLER DİZİNİ

Sayfa

Şekil 1.	5 durumlu bir Markov zinciri.....	10
Şekil 2.	Saklı yazı-tura deneylerinin sonuçlarını açıklayabilecek üç farklı Markov modelleri (a) 1 madeni paralı model (b) 2 madeni paralı model (c) 3 madeni paralı model.....	15
Şekil 3.	Kap ve top modeli	17
Şekil 4.	(a) 3 durumlu Markov modeli (b) 3 durumlu saklı Markov modeli.....	18
Şekil 5.	(a) $\alpha_t(i)$ ileri değişkeninin hesaplanabilmesi için gerekli işlemler dizisinin gösterilişi (b) t gözlemler ve i durumlar kafesi türünden $\alpha_t(i)$ nin hesaplanışı...	27
Şekil 6.	İleri ve geri değişkenlerinin gösterilişi.....	29
Şekil 7.	$\beta_T(i)$ geri değişkeninin hesaplanabilmesi için gerekli işlemler dizisinin gösterilişi	30
Şekil 8.	Sistemin t anında S_i ve $t+1$ anında ise S_j içinde bulunması bileşik olasılığının hesaplanması için gerekli işlemler dizisinin gösterilişi.....	35

1.GİRİŞ

Gerçek hayatta incelenen süreçler, çoğu zaman sinyal olarak karakterize edilebilecek gözlemlenebilir çıktılar oluşturur. Bu sinyaller, yapılarından dolayı kesikli nitelikte (örneğin; sonlu sayıda harften oluşan bir alfabadeki karakterler ya da kodkitabından bir niceliğin katsayılarının vektör olarak gösterilmesi) veya sürekli nitelikte (örneğin; konuşma örnekleri, sıcaklık ölçümleri, müzik gibi) de olabilirler. Sinyal kaynağı durağan (sinyal kaynağının istatistiksel özelliklerinin zamana göre değişmediği) ya da durağan olmayan (sinyal kaynağının istatistiksel özelliklerinin zamana göre değiştiği) yapıda olabilir. Burada sinyaller saf, katıksız (sımsızın tek bir sinyal kaynağından geldikleri durumda) olabileceği gibi, başka sinyal kaynakları (örneğin gürültü) veya iletim tahrifatları, yankılanımlar vb. tarafından bozunuma uğramış olabilir.

Araştırmalarda elde edilen bu sinyallerin, sinyal modelleriyle karakterize edilmesi, önemli bir problemdir. Sinyal modellerinin uygulanmasının sebepleri ile ilgili çeşitli nedenler vardır. Bunların ilki; sinyal sürecinde kullanılan, böylece istenilen özellikte çıktıları oluşturan bir sinyal modelinin teorik olarak tanımlanabileceği bir temel sağlayabilmesidir. Örneğin, gürültü ve iletim tahrifatları tarafından bozulmuş bir konuşma sinyalini güçlendirmek istediğimizde, sinyal modelini gürültüyü tamamen giderecek ve iletim tahrifatına yol açan işlemi kaldıracak bir sistem tasarlamak için kullanabiliriz. Sinyal modellerini önemli kılan ikinci bir neden ise, sahip oldukları potansiyel itibarıyla uygulamacıları sinyal kaynağını el altında bulundurmaya zorunda kalmaksızın, hakkında birçok şey öğrenebilme olanağı verebilecek niteliklere sahip olmalarıdır. Bu özellik, sinyallerin gerçek kaynaklarını elde etme maliyetleri yüksek olduğu durumlarda bilhassa önemlidir. Bu sayede, iyi bir sinyal modeliyle kaynaklar simüle edilebilir ve simülasyonlar yoluyla mümkün olduğunca bilgi edinilebilir. Son olarak, sinyal modellerini gerekli kılan önemli bir diğer neden, modellerin uygulamada mümkün olabilecek en iyi sonuçları vermesidir. Modellerin yardımıyla tahmin sistemleri, tanıma sistemleri ve teşhis sistemleri gibi önemli uygulama sistemleri etkili bir biçimde anlaşılır hale gelebilmektedir.

Verilen bir sinyalin özelliklerinin karakterize edilmesi için kullanılan birkaç olası sinyal modeli vardır. Sinyal modelleri genel olarak, deterministik modeller ve istatistiksel modeller olarak iki ana sınıfa ayrılabilir. Deterministik modellerde, genel olarak, sinyalin bir sinüs dalgası olması ya da üstel toplamlardan oluşması gibi bilinen bazı özgül sinyal özelliklerinden faydalanılmaktadır. Bu sınıftaki sinyal modelleri, genel olarak dolaysız teknik özelliklere sahiptir ve burada amaç, sinyal modellerine ait parametrelerin değerlerini belirlemektir (Örneğin sinüs dalgasının genliği, frekansı ve fazı ya da üstel genlikler ve

oranlar gibi). Sinyal modellerinin ikinci yaygın sınıfı, sinyallerin sadece istatistiksel özelliklerini tanımlamaya çalışan istatistiksel modeller kümesidir. Bu istatistik modellere örnek olarak Gauss süreçleri, Poisson süreçleri, Markov süreçleri, saklı Markov süreçleri gösterilebilir. İstatistiksel modellerin temelinde yatan varsayım, sinyalin rasgele süreç olarak iyi bir şekilde tanımlanabileceği ve stokastik süreçlerdeki parametrelerin kesin ve iyi tanımlanmış bir yöntemle belirlenebileceği (kestirilebileceği) olgusudur.

Önemli uygulamalardan olan konuşma süreçleri uygulamaları için, deterministik ve stokastik sinyal modellerinin ikisi de oldukça başarılıdır. Konu üzerinde yapılan çalışmaların tümünde bir stokastik sinyal modeli türü olan saklı Markov modeli (HMM) üzerinde durulmuştur. Bu modeller, haberleşme dilinde Markov kaynakları ya da Markov zinciri olasılık fonksiyonları olarak adlandırılmaktadır. Matematiksel yapısı zengin ve uygun bir şekilde kullanıldığında iyi sonuçlar veren Saklı Markov modelleri, 1960'ların sonu ve 70'lerin başında öne sürülmesine rağmen, konu ile ilgili çalışmaların yoğunlaşması ancak 90'lı yıllarda başlamıştır. Buradaki gecikmenin önemli iki nedeni, teorik yapının karmaşık olması ve bu modellerin uygulanmasındaki materyal sıkıntısıdır.

Saklı Markov modelleri, herbiri bir olasılık dağılımıyla (genellikle çok boyutlu) ilişkilendirilen sonlu durum kümeleridir. Durumlar arasındaki geçişler "geçiş olasılıkları" adı verilen olasılıklar kümesi ile belirlenir. Bir durumdan olasılık dağılımına göre sonuç veya gözlemler üretilebilir. Dışarıdaki gözlemci sadece sonucu görebilir, durumlar dışarıya "saklı"dır. Bu nedenle bu modeller Saklı Markov Modelleri olarak adlandırılmıştır.

Saklı Markov modelleri iki katmanlı rasgele bir süreç olarak düşünebilir. Alt katmanda durumları gözlenemeyen bir ayrık Markov zinciri, üst katmanda ise, bu zincirin her bir durumu için ayrı ayrı tanımlanmış gözlem olasılık yoğunlukları vardır.

Tez çalışmasının birinci bölümünde konuya bir giriş yapılmış; II. Bölümde olasılıksal süreçlerle ilgili bazı temel kavramlar tanıtılmıştır. Kesikli parametrelili Markov zincirleri Bölüm III' te kısaca verilmiş ve saklı Markov modeller için genelleştirilecek bazı basit örnekler incelenmiştir. Tez çalışmasının odağı olan saklı Markov modelleri Bölüm IV' te ele alınmıştır. Saklı Markov modellerindeki problemlerin incelenmesi ve çözümleri Bölüm V' te verilmiştir. Tez çalışmasının son bölümünde ise, sonuç ve tartışmalar yer almaktadır.

2. OLASILIKSAL SÜREÇLER

2.1. Giriş

Olasılıksal süreçler (stokastik süreçler) kuramı, olasılıksal kurallara göre, zaman içinde gelişen sistemleri inceler. Kuram, bilim ve tekniğin birçok dalında özellikle astronomi, biyoloji, endüstri, tıp gibi alanlarda başarıyla uygulanmakta; bilimsel öğrenme ve modelleme çalışmalarında etkin bir biçimde kullanılmaktadır.

Olasılıksal süreçlere basit bir örnek için bir evrenin büyüklüğünü göz önüne alalım. Başlangıçta ($t=0$ noktasında) evrenin büyüklüğü X_0 olabilir. Bu evrende sonraki gözleme dek, bazı bireyler evrenden ayrılabilir ya da bazı bireyler ortaya çıkabilir. Bu nedenle $t=1$ noktasında evrenin büyüklüğünü temsil eden X_1 , bir raslantı değişkenidir ve alacağı değer X_0 'a bağlıdır. $t=2$ noktasına ait olan X_2 değişkeni ise, olasılıksal olarak X_0 ve X_1 'e bağlıdır. Bu süreç için olasılıksal model, evrenin herhangi i ve j ($j > i$) zamanları arasındaki büyümesini (ya da azalmasını) gösterir ve X_0, X_1, \dots, X_i 'ye bağlıdır.

Kesikli zamanlı olasılıksal süreçlerde gözlemler yalnızca saptanmış zamanlarda yapılır. Sürekli zamanlı olasılıksal süreçlerde ise, gözlemler tüm olası zamanlarda düşünülür. Bu tür süreçlerin çözümlenmesi oldukça karmaşıktır.

2.2. Olasılıksal Sürecin Tanımı

Olasılıksal süreç, raslantı değişkenlerin oluşturdukları $\{X_t, t \in T\}$ kümesidir. Zaman, sürecin parametresidir. T 'ye indeks kümesi adı verilir ve x_t, X_t 'nin $t \in T$ zamanında aldığı değeri gösterir. $T=\{0, \pm 1, \pm 2, \dots\}$ ya da $T = \{0, 1, 2, \dots\}$ olduğunda sürece kesikli parametrelili süreç, $T = \{t: -\infty < t < \infty\}$ ya da $T = \{t: t \geq 0\}$ olduğunda sürekli parametrelili süreç denir (İnal, 1988).

Tanım bölgesinde her $h > 0$ $t \in T$ için $P(x-h < X_t < x+h)$ olasılığı varsa, x 'e $\{X_t, t \in T\}$ olasılıksal sürecin bir olası değeri ya da durumu adı verilir. Tüm olası durumların oluşturdukları kümeye de durum uzayı denir. Durum uzayında sonlu ya da sayılabilir sonsuzlukta durum varsa, sürecin kesikli durum uzaylı bir olasılıksal süreç olduğu söylenir. Durum uzayı kesikli ise incelenen süreçlere kesikli durum süreci; durum uzayı sürekli ise, ele alınan sürece sürekli durum süreci adı verilir. Tez çalışmasında durum uzayı S ile tanımlanacaktır.

2.2.1. Bağımsız artmalı süreç

$t_1 < t_2 < t_3 < t_4 < \dots$ için $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \dots$ bağımsız raslantı değişkenleri ise, $\{X_t, t \in T\}$ 'ye bağımsız artmalı süreç adı verilir. Böyle bir olasılıksal süreçte $X_{t+k} - X_t$ artması t 'den bağımsız ise, sürecin durağan bağımsız artmalı bir süreç olduğu söylenir (İnal, 1988).

2.2.2. Olasılıksal sürecin ortalama değer ve özdeğişke işlevleri

$\{X_t, t \in T\}$ olasılıksal sürecinde $E(X_t^2) < \infty$ olsun. $t \in T$ için,

$$m(t) = E(X_t)$$

biçiminde tanımlanan fonksiyona ortalama değer fonksiyonu ve $s, t \in T$ için,

$$K(s, t) = \text{Cov}(X_s, X_t)$$

biçiminde tanımlanan fonksiyona kovaryans fonksiyonu adı verilir. Kovaryans fonksiyonu $s, t \in T$ için simetriktir. Bu nedenle, $K(s, t) = K(t, s)$ eşitliği yazılabilir.

$V(X) = E(X^2) - [E(X)]^2$ ile $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ karşılaştırılırsa,

$$V(X_s) = \text{Cov}(X_s, X_s) = K(s, s) \quad (2.1)$$

elde edilir. Burada $V(X)$, X_s 'nin varyansıdır.

$\{X_t, t \in T\}$ olasılıksal sürecinde $K(s, t)$ yalnızca t ile s arasındaki farka bağlı ise, bu sürece kovaryansı durağan süreç adı verilir. Kovaryansı durağan süreçlerde, kovaryans fonksiyonu,

$$K(s, t) = K(t - s)$$

ya da t , iki zaman arasındaki farkı göstermek üzere,

$$\text{Cov}(X_s, X_{s+t}) = K(t) \quad (2.2)$$

olur. Eşitlik (2.1) ve (2.2)'den, her $s \in T$ için $V(X_s) = K(0)$ elde edilir.

Kovaryansı durağan bir sürecin ortalama değer fonksiyonunun değişmez olması gerekmez.

2.2.2. Durağan süreçler

Her ζ sayısı için,

$$Y_i = X_{t_i} + \zeta$$

biçiminde tanımlanan $\{Y_t, t \in T\}$ olasılıksal süreci. $\{X_t, t \in T\}$ olasılıksal süreci ile aynı ortalama değer ve kovaryans fonksiyonuna sahip ise, $\{X_t, t \in T\}$ 'ye durağan süreç adı verilir. Bu durumda ortalama değer fonksiyonu t 'den bağımsızdır ve kovaryans durağandır. Bu tanımlamaya göre ortalama değer fonksiyonu $m(t) = \mu'$ dir ve kovaryans fonksiyonu da $Cov(X_s, X_{s+t})=K(t)$ olarak yazılabilir. Burada söz edilen durağanlık tanımı uygulamalarda varlığının incelenmesi en kolay olan durağanlık tanımlamasıdır.

$(X_{t_1}, \dots, X_{t_k})$ rasgele vektörü ile $(Y_{t_1}, \dots, Y_{t_k})$ rasgele vektörü aynı dağılımlı ise, $\{X_t, t \in T\}$ 'nin k' inci dereceden durağan bir süreç olduğu söylenir. Burada verilen durağanlık tanımı kuvvetli anlamda bir durağanlık tanımıdır. Şayet $\{X_t, t \in T\}$ olasılıksal süreci her k için k' inci dereceden durağan süreç ise, bu sürece de k' inci dereceden kovaryans durağan süreci adı verilir (İnal,1988).

2.3. Markov Sürecinin Tanımı

İndeks kümesinde yer alan $t_1 < t_2 < \dots < t_n$ zaman noktaları için, X_{t_n} 'nin $X_{t_1}, X_{t_2}, X_{t_3}, \dots, X_{t_{n-1}}$ 'lerin aldığı değerlere koşullu dağılımı yalnızca $X_{t_{n-1}}$ 'in değerine bağlı ise, $\{X_t, t \in T\}$ sürecine bir Markov süreci adı verilir. Buna göre, herhangi gerçel x_1, \dots, x_n sayıları için,

$$P(X_{t_n}=x_n/X_{t_1}=x_1, X_{t_2}=x_2, \dots, X_{t_{n-1}}=x_{n-1})= P(X_{t_n}=x_n/X_{t_{n-1}}=x_{n-1}) \quad (2.3)$$

tanımı söz konusudur. Eşitlik (2.3)'e Markov özelliği adı verilir. Yukarıda tanımlanan eşitliğin sözel ifadesi ise, sürecin şu anki durumu sadece bir önceki duruma bağlıdır ve geçmiş t_1, \dots, t_{n-2} 'deki durumlardan bağımsızdır (İnal,1988).

Markov süreçleri, sürecin parametresi olan zamanın ve durum uzayının niteliğine göre sınıflandırılır. Durum uzayı kesikli olan Markov sürecine Markov zinciri adı verilir. Uygulamalarda $\{0,1,2,\dots\}$ tamsayılar kümesi, Markov zincirlerinin durum uzaylarının tanımlanmasında sık sık kullanılır. Durum uzayı sonlu ise, zincire sonlu Markov zinciri, aksi durumda incelenen sürece sonsuz Markov zinciri adı verilir. Markov süreçlerinin durum uzayına ve zamana bağlı olarak sınıflandırılması Tablo 2.1' de verilmiştir.

		S	
		Kesikli	Sürekli
T	Kesikli	Kesikli Değişirgeli Markov Zinciri	Kesikli Değişirgeli Markov Süreci
	Sürekli	Sürekli Değişirgeli Markov Zinciri	Kesikli Değişirgeli Markov Süreci

Tablo 2.1. Markov Süreçlerinin Sınıflandırılması

Tez çalışmamızda amaç saklı Markov modellerinin incelenmesi olduğundan öncelikle kesikli parametrelili Markov zincirleri izleyen bölümde kısaca ele alınacaktır.

3. KESİKLİ PARAMETRELİ MARKOV ZİNCİRLERİ

3.1. Geçiş Olasılıkları

$\{X_t, t=0,1,2,\dots\}$ kesikli parametrelili bir Markov zinciri olsun. Kesikli parametrelili olduğundan, bu zincir $\{X_0, X_1, X_2,\dots\}$ biçiminde gösterilebilir. Durum uzayındaki i ve j durumları için tanımlanan,

$$p_{ij}(m,n)=P(X_{m+n}=j/X_m=i) , \quad i,j \in S \text{ için} \quad (3.1)$$

koşullu olasılığı, geçiş olasılığı fonksiyonu adını alır. Eşitlik (3.1)'de verilen olasılık yalnızca n zaman aralığına bağlı, m zamanına bağlı değilse, zincire zaman içinde homojen Markov zinciri ya da durağan geçiş olasılıklı Markov zinciri adı verilir (İnal,1988). Bu tür zincirlerde,

$$p_{ij}(n) = P(X_{m+n}=j/X_m=i) , \quad i, j \in S \text{ için} \quad (3.2)$$

olasılığın n-adım geçiş olasılığı denir. Eşitlik (3.2)'de n=1 ise $p_{ij}(1)$, p_{ij} biçiminde gösterilir ve

$$p_{ij} = P(X_{m+1}=j/X_m=i) , \quad i, j \in S \text{ için} \quad (3.3)$$

olasılığı bir-adım geçiş olasılığı adı verilir.

Herhangi bir i durumundan durum uzayındaki diğer tüm durumlara i durumu da dahil olmak üzere geçilebildiğinden standart olasılık tanımları çerçevesinde

$$\sum_j p_{ij}^{(n)} = 1 , \quad \text{her } i \in S ; n=0,1,2,\dots \text{ için} \quad (3.4)$$

yazılabilir ve n=0 için,

$$p_{ij}(0) = 1, \quad i=j \text{ için}$$

$$= 0, \quad i \neq j \text{ için}$$

olur.

3.2. Olasılık Vektörü Ve Olasılık Matrisi

Bir $u = [u_1 \ u_2 \ \dots \ u_n]$ vektörünün olasılık vektörü olabilmesi için, aşağıda verilen iki koşulu sağlaması gerekir. Bunlar,

- 1) Olasılık vektöründe negatif üyelerin olmaması ve
- 2) Öğelerin toplamının 1'e eşit olmasıdır.

$A = [a_{ij}]$ kare matrisinin tüm satırları olasılık vektörü ise, böyle matrislere olasılık matrisi adı verilir.

TEOREM 1: $A=[a_{ij}]$ matrisi $n \times n$ boyutlu olasılık matrisi ve u vektörü n öğeli olasılık vektörü ise, uA , bir olasılık vektörüdür.

TANIT: uA vektörü,

$$u \ A = [u_1 \dots u_n] \begin{bmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1n} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ a_{n1} & \cdot & \cdot & \cdot & a_{nn} \end{bmatrix}$$

$$= [u_1 a_{11} + \dots + u_n a_{n1} + \dots + u_1 a_{1n} + \dots + u_n a_{nn}]$$

olur. Vektörün öğeleri pozitifdir ve öğelerin toplamı 1'e eşittir.

$$u_1 a_{11} + \dots + u_n a_{n1} + \dots + u_1 a_{1n} + \dots + u_n a_{nn} = u_1 (a_{11} + \dots + a_{1n}) + \dots + u_n (a_{n1} + \dots + a_{nn})$$

$$= u_1 + \dots + u_n = 1$$

TEOREM 2: $A=[a_{ij}]$ ve $B=[b_{ij}]$, eşit boyutlu olasılık matrisleri ise, AB olasılık matrisidir.

TANIT: AB matrisini,

$$AB = \begin{bmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1n} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ a_{n1} & \cdot & \cdot & \cdot & a_{nn} \end{bmatrix} \begin{bmatrix} b_{11} & \cdot & \cdot & \cdot & b_{1n} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ b_{n1} & \cdot & \cdot & \cdot & b_{nn} \end{bmatrix}$$

biçiminde yazılabilir. A'nın her satırı olasılık vektörüdür. AB'nin satırları A'nın satırları ile B matrisini çarparak elde edildiği için Teorem 1'e göre AB'nin satırları olasılık vektörü ve AB, olasılık matrisidir.

TEOREM 3: $A=[a_{ij}]$ olasılık matrisi ise, A^n de (tüm pozitif n değerler için) bir olasılık matrisidir.

TANIT: Teorem 2'de B yerine A alınırsa A^2 olasılık matrisi olur. Bu yol izlendiğinde, A^3, A^4, \dots matrislerinin de olasılık matrisleri oldukları görülür.

3.3. Geçiş Matrisi

Durum uzayı $S=\{0,1,2,\dots\}$ olan bir Markov zincirinde, bir adım geçiş olasılıklarının oluşturdukları matrise geçiş matrisi denir ve aşağıdaki şekilde tanımlanır. Buradaki tanımda durum uzayının sayılabilir sonsuzlukta olduğu kabul edilmiştir.

$$P = [p_{ij}] = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & \cdot & \cdot & \cdot \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ \cdot \\ \cdot \\ \cdot \end{matrix} & \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdot & \cdot & \cdot \\ p_{10} & p_{11} & p_{12} & \cdot & \cdot & \cdot \\ p_{20} & p_{21} & p_{22} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \end{matrix}$$

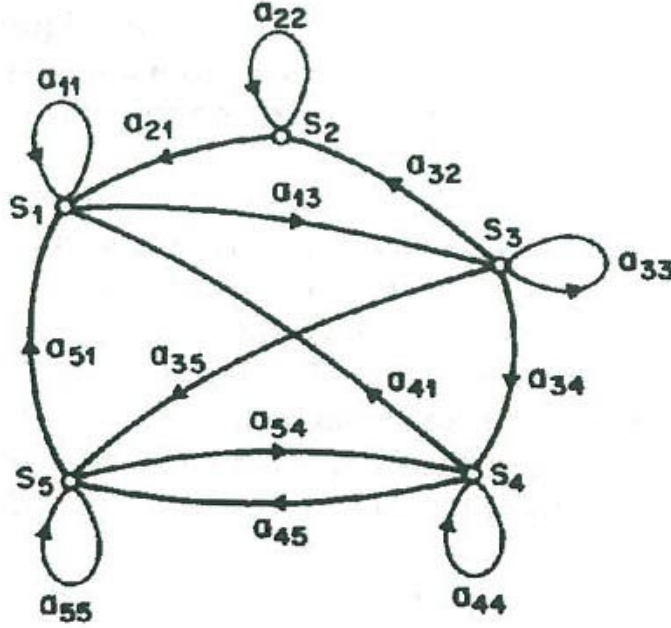
burada yukarıda tanımlanan geçiş matrisi aşağıdaki koşulları sağlar (İnal,1988).

1) $p_{ij} \geq 0$, tüm $i, j \in S$ için,

2) $\sum_j p_{ij} = 1$, tüm $i \in S$ için

3.4. Kesikli Markov Süreci

Bu sürecin daha açık bir biçimde ifade edilmesi için Şekil 1'de sunulan örnek ele alınmıştır. Burada durum sayısı 5 olarak alınmıştır. Durumlardan durumlara geçişlerin söz konusu olduğu bu örnek için mevcut olası tüm kombinasyonlar aşağıda görülmektedir.



Şekil 1. 5 durumlu bir Markov zinciri

Burada süreç; düzenli aralıklara sahip kesikli zamanlarda, durumlarla ilgili olasılıklar kümesine göre bir durum değişikliğine (aynı duruma dönmesi olasıdır) uğramaktadır. Durum değişikliklerinin meydana geldiği anlar $t=1,2,\dots$ olarak tanımlanmakta ve t anındaki gerçek durum q_t olarak gösterilmektedir. Yukarıdaki sisteme ait tam bir olasılık tanımı, genel olarak mevcut durum (t anındaki) ile bundan önceki durumların özelliklerinin belirtilmesini gerektirecektir. Özel olarak, kesikli birinci dereceden Markov zinciri için, bu olasılıkların tanımı, sadece mevcut duruma ve daha önceki durumlara göre kesilerek kısaltılacaktır. Buna göre,

$$P[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \dots] = P[q_t = S_j | q_{t-1} = S_i] \quad (3.5)$$

yazılabilir. Burada (3.5)'deki eşitliğin sağ tarafının geçmiş zamandan bağımsız olduğu söylenebilir. Bu olasılıklar aşağıdaki geçiş olasılıkları olarak ifade edilebilir.

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i] , \quad 1 \leq i, j \leq N \quad (3.6)$$

Burada standart olasılık kuralları çerçevesinde

$$a_{ij} \geq 0 \quad (3.7a)$$

$$\sum_{j=1}^N a_{ij} = 1 \quad (3.7b)$$

kısıtların varlığı söz konusudur (Rabiner, 1989).

Yukarıdaki tahmini süreç, sürecin çıktısı, her bir durum gözlemlenebilir bir olaya karşılık gelmek üzere zamanın her bir anındaki durumların kümesi olduğundan, gözlemlenebilir bir Markov modeli olarak adlandırılabilir. Kesikli Markov zinciri ile ilgili anlaşılır bir örnek olarak burada 3 durumlu bir Markov zinciri örnek olarak verilecektir. Gün içinde herhangi bir zamanda havanın aşağıda verilen durumlardan biri olarak gözlemlendiği varsayalım.

Durum 1: yağmurlu ya da karlı

Durum 2: bulutlu

Durum 3: güneşli

Mantıksal olarak herhangi bir t gününde havanın, yukarıdaki durumdan sadece biri ile tanımlanacağı açıktır. Burada bir adım geçiş olasılıklarının yer aldığı geçiş matrisi aşağıdaki şekilde tanımlandığı kabul edilsin.

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

Araştırmacı birinci gündeki (t=1) havanın güneşli (Durum 3) olduğunu varsayarak şu soruyu sorabilir: "Sonraki 7 gün için havanın "güneşli-güneşli-yağmurlu-yağmurlu- güneşli-bulutlu- güneşli" olması olasılığı nedir?".Bu soruyu biçimsel olarak ifade etmek için, t=1,2,3,...,8' e karşılık gelmek üzere O gözlem dizisi $O = \{S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3\}$ biçiminde

tanımlanabilir ve bundan faydalanarak, herhangi bir model için O vektörünün gerçekleşme olasılığının elde edilmesi amaçlanabilir. Bu olasılık, ilk durum olasılığı bilindiği için,

$$\pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N \quad (3.8)$$

$$\begin{aligned} P(O|Model) &= P[S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3 | Model] \\ &= P[S_3] \cdot P[S_3|S_3] \cdot P[S_3|S_3] \cdot P[S_1|S_3] \cdot P[S_1|S_1] \\ &\quad \cdot P[S_3|S_1] \cdot P[S_2|S_3] \cdot P[S_3|S_2] \\ &= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\ &= 1 \cdot (0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2) \\ &= 1.536 \times 10^{-4} \end{aligned}$$

olarak ifade edilebilir ve hesaplanabilir.

Burada sorabilecek ve modeli kullanarak cevaplanabilecek bir diğer ilginç soru ise, modelin bilinen bir durumda olduğunu varsayıldığında, tam olarak d kadar gün boyunca havanın aynı durumda kalması olasılığının ne olacağıdır. Bu olasılık, verilen model için aşağıdaki şekilde tanımlanabilir.

$$O = \left\{ S_1, S_2, S_3, \dots, S_d, S_{d+1} \neq S_d \right\}$$

Bu olasılık Eşitlik (3.9) ile ifade edilebilir.

$$P(O|Model, q_1 = S_i) = (a_{ij})^{d-1} (1 - a_{ij}) = p_i(d) \quad (3.9)$$

Burada $p_i(d)$ 'nin büyüklüğü i durumundaki d süre için kesikli olasılık yoğunluk fonksiyonunun değeridir. Bu üstel süre yoğunluğu, bir Markov zincirindeki durumun süresinin karakteristiğidir. $p_i(d)$ 'den faydalanarak, söz konusu durum için

$$\bar{d}_i = \sum_{d=1}^{\infty} d p_i(d) \quad (3.10a)$$

$$= \sum_{d=1}^{\infty} d(a_{ii})^{d-1}(1-a_{ii}) = \frac{1}{1-a_{ii}} \quad (3.10b)$$

biçiminde başlamak koşuluyla, beklenen gözlem (süre) miktarını hesaplanabilir. Böylece, güneşli olarak geçmesi beklenen sonraki günlerin sayısı modele göre,

$$\bar{d}_i = \frac{1}{1-a_{ii}} = \frac{1}{1-0.8} = \frac{1}{0.2} = 5$$

dir. Bulutlu günlerin sayısı,

$$\bar{d}_i = \frac{1}{1-a_{ii}} = \frac{1}{1-0.6} = \frac{1}{0.4} = 2.5$$

dir ve yağmurlu günlerin sayısı ise

$$\bar{d}_i = \frac{1}{1-a_{ii}} = \frac{1}{1-0.4} = \frac{1}{0.6} = 1.67$$

olarak hesaplanır.

4. SAKLI MARKOV MODELLERİ

Tez çalışmasında bu noktaya kadar, her bir durumun gözlemlenebilir bir olaya karşılık geldiği Markov modelleri ele alınmıştır. Ancak bu modeller, birçok önemli problemin çözümünde uygulanabilirlik açısından son derece kısıtlayıcı modellerdir. Bu bölümde ise, Markov modellerinde mevcut olan kısıtlayıcı durumları ortadan kaldıran saklı Markov modelleri ele alınacaktır. Bu modellerde gözlemlerin elde edildiği durumlar gözlenemediğinden ele alınan süreç gizli bir süreçtir ve ancak gözlem dizilerini oluşturacak başka bir olasılıksal süreç kümesi vasıtasıyla gözlemlenebilecek temel bir olasılıksal sürece sahip iki kat daha derinde saklanan Markov modellerini de içerir. Saklı Markov modellerinin daha anlaşılır bir şekilde ifade edilmesi için aşağıda verilen yazı-tura ve kap-top modelleri ele alınacaktır.

Yazı Tura Modelleri

Burada içinde neler olduğunun görülmesini engelleyen bir bariyer olan bir odanın var olduğu düşünölsün. Bariyerin bir tarafında, bir ya da birden çok madeni para ile yazı tura deneyleri yapan bir birey olsun. Bu kişi, tam olarak ne yaptığı konusunda bariyerin diğery tarafında olan kişiyi bilgilendirmesin ve sadece havaya atılan her bir madeni paranın sonucunu söylesin. Bu tür bir denemede araştırmacının elinde sadece yazı ve turalardan oluşan bir gözlem dizisi olacaktır. Burada elde edilen gözlem dizisi, K yazıyı ve S de turayı ifade etmek üzere şu şekilde olacaktır.

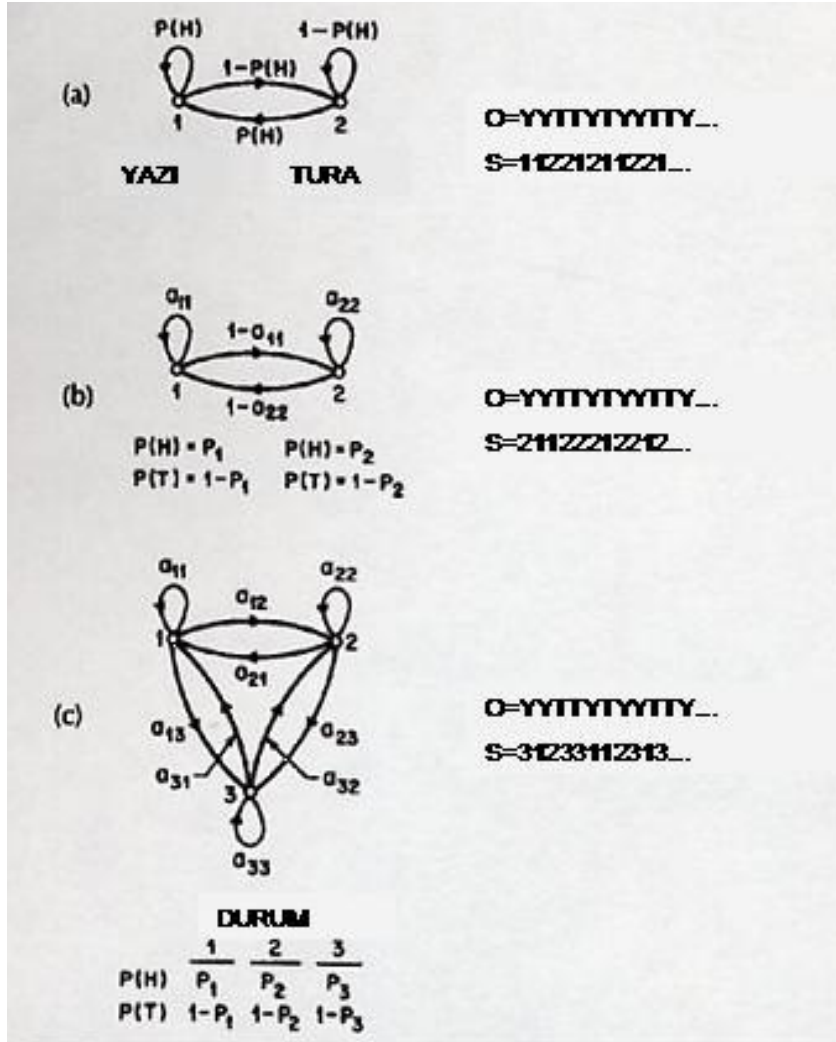
$$\begin{aligned} O &= O_1 O_2 O_3 \dots O_T \\ &= KKSSSKSSK\dots K \end{aligned}$$

Yukarıdaki senaryo göz önünde alındığında, önemli olan nokta, gözlemlenen yazı ve turalar dizisini açıklayacak bir saklı Markov modelini nasıl yapılandırılacağıdır. Burada karşılaşılabacak ilk soru, modeldeki durumların neye karşılık geleceğı ve bundan sonraki soru da modelin içinde kaç tane durumun olması gerektiğine karar verilmesidir. Olası bir seçenek, bu denemelerin sadece tek taraflı bir madeni parayla yapılmış olabileceğinin düşünölmesidir. Böyle bir varsayımda, durumların her biri madeni paranın bir yüzüne karşılık gelecek biçimde iki durumlu bir Markov modeli ile karşılaşılır. Bu model Şekil 2(a)'da resimle gösterilmiştir (Rabiner,1989). Bu durumda Markov modeli gözlemlenebilir olup, modelin özelliklerinin tam olarak belirtilebilmesine ilişkin tek sorun, paranın değışmiş yansızlığı (yani; örneğın tura gelmesi olasılığı) için en iyi değeryin ne olacağı hakkında kararın verilmesi olacaktır. Şekil.2(a)'dakine benzer bir saklı Markov modeli, durumun tek bir biçimde koşullandırılmış madeni paraya denk geldiğı ve bilinmeyen parametrenin madeni paranın yanlılık değeryi olduğu bozulmuş bir 1-durumlu model olacaktır. Burada durumun bozulmuş olma nedeni madeni paranın hileli bir yapıya sahip olmasıdır.

Havaya atılan madeni paranın sonucu ile ilgili olarak gözlemlenen diziyi açıklayacak ikinci bir saklı Markov modeli formu da Şekil.2.(b)'de verilmiştir. Buradaki modelde iki durum bulunmakta olup, her bir durum havaya atılan farklı ve yanlı bir madeni paraya karşılık gelmektedir. Durumların her biri, yazı ve turalardan oluşan bir olasılık dağılımı ile tanımlanırken; durumlar arasındaki geçişler de bir geçiş matrisi ile açıklanır. Durumlardan

durumlara geçişi belirleyecek fiziksel mekanizma ise bağımsız para atma olaylarının bir kümesi ya da diğer bazı olasılık olayları ile olabilecektir.

Havaya atılan madeni paranın sonucu ile ilgili olarak gözlemlenen diziyi açıklayacak üçüncü bir saklı Markov modeli formu ise, Şekil.2.(c)'de verilmiştir. Bu model, üç koşullandırılmış madeni paradan faydalanılmasına ve bu üçü arasında bazı olasılık olaylarına dayanarak seçim yapılmasına karşılık gelmektedir.



Şekil 2. Saklı yazı- tura deneylerinin sonuçlarını açıklayabilecek üç farklı Markov modelleri

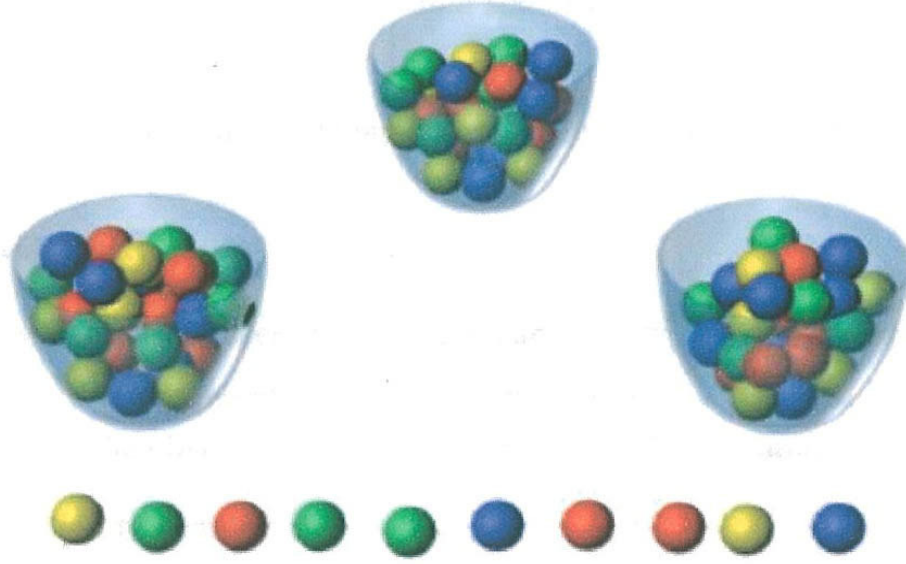
- (a) 1 madeni paralı model
- (b) 2 madeni paralı model
- (c) 3 madeni paralı model

Yukarıdaki şekilde yer alan birinci model, tek bir madeni para içeren modeli, ikinci model madeni para sayısının iki adet olduğu saklı Markov modelini ve son model ise kullanılan madeni para sayısının üç adet olduğu durumu içermektedir. Burada dikkat edilmesi gereken husus madeni para sayısı arttıkça modellerin daha karmaşık bir hal alacağıdır. Yazı-tura modelleri saklı Markov modellerinin biçiminin daha iyi anlaşılabilmesi için farklı madeni para sayılarında da incelenebilir. Burada gözlemlenen yazı tura dizilerinin açıklanmasına ilişkin Şekil.2'de gösterilen üç model arasından yapılmış bir seçim için, doğal olarak sorulması gereken soru, hangi modelin gerçek gözlemleri en iyi biçimde karşılayacağıdır. Şekil.2(a)'daki basit bir madeni paralı modelde bilinmeyen parametre sayısı 1, Şekil.2 (b)'deki iki madeni paralı modelde bilinmeyen parametre sayısı 4 ve Şekil.2(c) 'deki üç madeni paralı modelde bilinmeyen parametre sayısı 9 olacaktır. Bu nedenle, daha büyük saklı Markov modelleri, zorluğuna karşın daha fazla serbestlik sağlayacağından, bir yazı tura serisinin modellenmesinde, kendilerine denk düşen daha küçük modellere göre, yapılarında mevcut olan özelliklerinden dolayı daha elverişli görünmektedir.

Kap ve Top Modeli

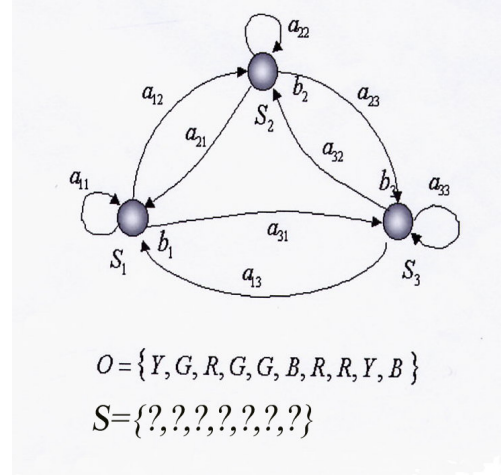
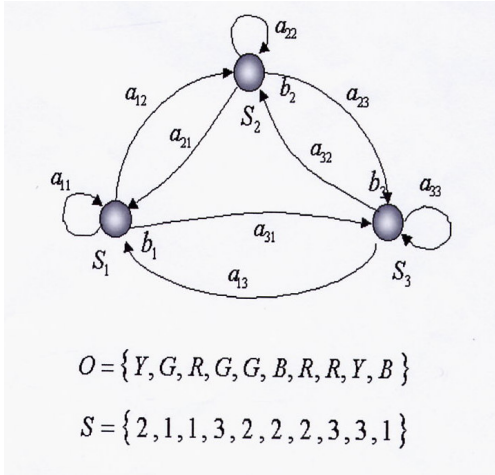
Bu çalışmada saklı Markov modeli teorisindeki fikirleri daha karmaşık durumlara uygun gelecek biçimde genişletmek için, yukarıda açıklanan yazı-tura modeline ek olarak Şekil.3'te tanımlanan kap ve top sistemi verilecektir. Bu modelde bir odanın içinde N sayıda içinde renkli toplar olan camdan kaplar ve seçimi yapan bir kişinin olduğu varsayılmıştır. Burada her bir kabın içinde, M farklı renkte çok sayıda top bulunmaktadır. Odadaki kişi rasgele olarak herhangi bir kabı seçmekte ve seçilen bu kaptan rasgele bir top almaktadır. Kişi topun rengini gözlem olarak kaydedilmek üzere seçimi görmeyen diğer kişiye beyan ettikten sonra çektiği topu yine aynı kaba geri koymaktadır. Bu işlem mevcut kapla ilgili rasgele seçim sürecine göre yeni bir kap seçilerek ve rasgele bir top alınarak yinelenmektedir.

Bu sürecin tamamı, gözlemlenebilir saklı Markov modeli çıktısı olarak modellendirmek istediğimiz sonlu bir renk gözlemleri dizisi üretecektir. Şurası açık olmalıdır ki, kap ve top sürecine karşılık gelen en basit saklı Markov modeli, her durumun özel bir kaba karşılık geldiği ve her bir durum için bir top rengi olasılığının tanımlanabileceği modeldir. Kapların seçimi, saklı Markov modeline ait durum geçiş matrisine göre belirlenecektir.



Şekil 3. Kap ve top modeli

Yukarıda tanımlanan kap ve top modelinin durumları bilindiğinde Bölüm 3'te tanımlanan kesikli parametrelili Markov zinciri elde edilir. Bu model aşağıda verilen Şekil 4(a)'da tanımlanmaktadır. Burada hem gözlemler hem de gözlemleri yaratan durumlar (kaplar) bilinmektedir. Kap ve top modelinde 10 deneme sonucunda örnek olarak elde edilen gözlem dizisinin aşağıda verildiği gibi $O = \{\text{sarı, yeşil, kırmızı, yeşil, yeşil, mavi, kırmızı, kırmızı, sarı, mavi}\}$ olduğu görülmekte ve bu topların sırasıyla 2,1,1,3,2,2,2,3,3,1 nolu kaplardan çekildikleri görülmektedir. Buna karşın Şekil 4(b)'de tanımlanan modelde $O = \{\text{sarı, yeşil, kırmızı, yeşil, yeşil, mavi, kırmızı, kırmızı, sarı, mavi}\}$ şeklinde elde edilen gözlemler dizisi için durumlar gizlidir. Burada örneğin birinci denemede elde edilen sarı topun mevcut üç kapın hangisinden geldiği bilinmemektedir. Bu durum saklı Markov modeli ile Markov modeli arasındaki temel farklılıktır. Gerçek hayatta bu nedenle saklı Markov modellerinin daha geniş bir kullanım alanı olacağı açıktır. Özellikle kaynağı belli olmayan sinyallerin modellenmesinde saklı Markov modellerinin daha başarılı olacağı açıkça görülmektedir.



Şekil 4. (a) 3 durumlu Markov modeli
(b) 3 durumlu saklı Markov Modeli

4.1. Saklı Markov Modelini Oluşturan Unsurlar

Yukarıda verilen örnekler, saklı Markov modelinin ne olduğu ve basit senaryolara nasıl uygulanabileceği konusunda oldukça iyi bir fikir vermektedir. Çalışmanın bu bölümünde saklı Markov modelini oluşturan unsurlar biçimsel olarak tanımlanacak ve modelin gözlem dizilerini nasıl ürettiği açıklanacaktır.

Bir saklı Markov modeli aşağıdaki biçimde tanımlanabilir:

1) N , model içindeki durumların sayısıdır. Durumlar gizli olsalar da, bir çok pratik uygulama için çoğu zaman durumlara ya da modeldeki durumlar kümesine ilişkin bazı belirgin fiziksel anlamlar mevcuttur. Bundan dolayı, örneğin yazı tura deneylerinde, her durum ayrı bir koşullandırılmış paraya karşılık gelir. Kap ve top modelinde, durumlar kaplara karşılık gelir. Genel olarak, durumlar kendi içlerinde, herhangi bir duruma herhangi başka bir durumdan ulaşılacağı biçimde bağlantılıdır. Bununla birlikte, konunun ilerleyen bölümlerinde, durumların olası diğer iç bağlantılarının da sıklıkla önemli olduğuna değinilecektir. Çalışmada durum kümesi $S = \{S_1, S_2, \dots, S_N\}$ biçiminde ve t anındaki durum ise q_t olarak gösterilecektir.

2) M, her bir duruma ait ayrık gözlem sembollerinin sayısıdır. Şayet gözlemler sürekli ise, M sonsuzdur. Gözlem sembolleri, modellenen sistemin fiziksel çıktısına karşılık gelir. Yazı tura deneyleri için, gözlemlerin sembolleri sadece yazı ve turadan ibaret olup; kap ve top modelinde ise kaplardan seçilen topların renkleri bu sembollerin yerini almıştır. Çalışmada sembollerin her biri $V = \{v_1, v_2, \dots, v_M\}$ biçiminde gösterilecektir.

$$3) \quad a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N \quad (4.1)$$

olmak üzere, durum geçiş olasılık dağılımı $A = \{a_{ij}\}$ 'dir.

Herhangi bir durumun tek bir adımda herhangi başka bir duruma dönüşebileceği özel bir durum için, her i,j için $a_{ij} > 0$ olacaktır. Diğer saklı Markov modeli türleri için, bir ya da birden fazla (i,j) çifti için $a_{ij} = 0$ olacaktır. Geçiş olasılıkları standart olasılık kurallarını sağlamalıdır (Warakagoda, 1996).

$$a_{ij} \geq 0, \quad 1 \leq i, j \leq N$$

ve

$$\sum_{j=1}^N a_{ij} = 1, \quad 1 \leq i \leq N$$

$$4) \quad b_j(k) = P[v_k \text{ at } t | q_t = S_j], \quad 1 \leq j \leq N \quad (4.2)$$

$$1 \leq k \leq M$$

olmak üzere, j durumundaki gözlem sembolü olasılığı dağılımı $B = \{b_j(k)\}$ 'dir.

$b_j(k)$ için aşağıdaki stokastik kısıtlar sağlanmalıdır. (Warakagoda, 1996)

$$b_{ij} \geq 0, \quad 1 \leq j \leq N, \quad 1 \leq k \leq M$$

ve

$$\sum_{k=1}^M b_j(k) = 1, \quad 1 \leq j \leq N$$

$$5) \quad \pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N \quad (4.3)$$

olmak üzere başlangıç durum dağılımı $\pi = \{\pi_i\}$, dir.

N, M, A, B ve π 'nin uygun değerleri için (her bir O_t gözlemi V_t sembollerinden biri ve T ise dizideki gözlemlerin sayısı olmak üzere) saklı Markov modeli,

$$O = O_1 O_2 \dots O_T \quad (4.4)$$

gibi bir gözlem dizisini verecek bir üreteç olarak kullanılabilir.

Bunun için;

Adım 1) ilk durum dağılımı π 'ye bağlı olarak, $q_1 = S_i$ olacak biçimde başlangıç durumu seçilir.

Adım 2) $t=1$ olarak alınır.

Adım 3) S_i durumundaki sembol olasılığı dağılımına yani $b_i(k)$ 'ya göre $O_t = v_k$ olarak seçim yapılır.

Adım 4) S_i durumunun durum geçiş olasılığı dağılımının, yani a_{ij} 'ye göre yeni bir $q_{t+1} = S_j$ durumuna geçilir.

Adım 5) $t=t+1$ olarak alınır; $t < T$ ise 3. adıma geri dönülür. Diğer durumlarda ise işlem sonlandırılır.

Yukarıdaki işlemler, hem bir gözlem üretici olarak ve hem de göz önünde bulundurulmuş bir gözlem dizisinin, uygun bir saklı Markov modeli tarafından nasıl üretildiğinin bir modeli olarak kullanılabilir.

Yukarıdaki değerlendirmelerden, saklı bir Markov modelinin özelliklerinin tam olarak belirlenebilmesi için, model parametreleri olan N ve M'nin, gözlem sembollerinin ve üç olasılık ölçümünün A, B ve π 'nin belirtilmesinin gerekli olduğu görülür. Gösterimde kolaylık olması açısından, modeldeki parametreler kümesinin tamamını kısaca

$$\lambda = (A, B, \pi) \quad (4.5)$$

biçiminde gösterilecektir (Warakagoda,1996).

4.2. Saklı Markov Modelinin Varsayımları

Matematiksel ve hesaba dayalı işlemlerde kolaylık sağlaması açısından saklı Markov modelinin teorisinde aşağıdaki varsayımlar kullanılmaktadır. Bu varsayımlar,

- 1) Markov varsayımı,
- 2) durağanlık varsayımı ve
- 3) çıktı bağımsızlık varsayımdır.

Burada yukarıda belirtilen varsayımlar kısaca açıklanacaktır.

1) Markov varsayımı:

Saklı Markov modelinin tanımında verildiği üzere geçiş olasılıkları aşağıdaki gibi tanımlanmıştır.

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i]$$

Kesim 2.3'te verildiği gibi burada da gelecek durumun yalnızca şimdiki duruma bağlı olduğu varsayılmaktadır. Burada model gerçekte birinci derece saklı Markov modeli olur. Gelecek durum geçmişteki k duruma dayanabilirse, ki bu durum bazı modellerde geçerli olur, bu durumda model k. derece saklı Markov modeli olarak adlandırılır. Geçiş olasılıkları ise aşağıdaki gibi olur.

$$a_{i_1 i_2 \dots i_k j} = P[q_{t+1} = S_j | q_t = S_{i_1}, q_{t-1} = S_{i_2}, \dots, q_{t-k+1} = S_{i_k}] , 1 \leq i_1, i_2, \dots, i_k, j \leq N$$

Yüksek dereceli saklı Markov modelleri yapı itibariyle çok karmaşıktır. Bu nedenle birinci derece saklı Markov modelleri çok yaygındır. Ancak son yıllarda araştırmacılar yüksek dereceli saklı Markov modellerinin kullanımı üzerine çalışmaktadırlar.

2) Durağanlık varsayımı

Bu varsayımda geçişlerin yer aldığı durum geçiş olasılıklarının geçerli durumdan bağımsız oldukları varsayılmaktadır. Matematiksel olarak aşağıdaki biçimde tanımlanır.

$$P[q_{t_1+1} = S_j | q_{t_1} = S_i] = P[q_{t_2+1} = S_j | q_{t_2} = S_i] \quad \text{her } t_1 \text{ ve } t_2 \text{ için}$$

3) Çıktı bağımsızlık varsayımı

Bu varsayım ise var olan, geçerli çıktının (gözlem) önceki gözlemlerden istatistiksel olarak bağımsız olması gerektiğini vurgular. Bu varsayımı, $O = O_1, O_2, \dots, O_t$ gözlem serisi ele alarak matematiksel olarak formüle edebiliriz.

Saklı Markov modeli λ için varsayıma göre aşağıdaki biçimde tanımlanır.

$$P[O | q_1, q_2, \dots, q_T, \lambda] = \prod_{t=1}^T P(o_t | q_t, \lambda)$$

Diğer iki varsayımdan farklı olarak, bu varsayımın geçerliliği çok sınırlıdır. Bazı hallerde bu varsayım yeterince uygun olmayabilir. Böylece saklı Markov modellerinin kuvvetli zayıflığı ortaya çıkar. (Warakagoda, 1996)

5. SAKLI MARKOV MODELLERİNİN ÜÇ TEMEL PROBLEMİ

Bir önceki bölümdeki saklı Markov modeli biçimi göz önüne alındığında, modelin gerçek verilere uygulanabilmesi için çözülmesi gereken önemli üç ana problemin olduğu göze çarpmaktadır. Problemler aşağıda sunulmuştur:

Problem 1: Bu problem literatürde değerlendirme problemi olarak adlandırılmaktadır. Burada vurgulanan problem " $O = O_1, O_2, \dots, O_t$ bir gözlem dizisi ve $\lambda = (A, B, \pi)$ ise bir model olmak üzere, bu modele ait gözlem dizisi olasılığı $P(O|\lambda)$ etkin bir biçimde nasıl hesaplanabilir?" sorusunun nasıl cevaplanacağıdır.

Problem 2: Literatürde bu probleme kodlama problemi olarak yer almıştır. Bu problem ile " $O = O_1, O_2, \dots, O_t$ bir gözlem dizisi ve $\lambda = (A, B, \pi)$ ise bir model olmak üzere, modele uygun olarak gözlemleri en iyi açıklayabilen ve $Q = q_1, q_2, \dots, q_T$ şeklinde gösterilen durum dizisi nasıl seçilebilir?" sorusunun cevaplanması amaçlanmaktadır.

Problem 3: Öğrenme problemi olarak da adlandırılan bu problemde ise “ $P(O|\lambda)$ olasılığının maksimize edilebilmesi için $\lambda=(A,B,\pi)$ model parametrelerinin nasıl ayarlanması gereklidir?” sorusuna yanıt aranmaktadır.

1. problem, söz konusu bir model ve gözlemler dizisi için, gözlemlenen dizinin model tarafından üretilme olasılığının nasıl hesaplanacağı ile ilgili bir problemdir. Ayrıca, bu problem, belirli bir modelin belirli bir gözlem dizisi ile ne kadar eşleştireceği problemi olarak da düşünülebilir. Bu bakış açısı son derece kullanışlıdır. Örneğin, birbirine rakip iki model arasından seçim yapılması istenildiğinde, 1. problemin çözümü gözlemlere en uygun olan modeli seçme şansını verecektir.

2. problem, modelin gizli kalan kısmını açığa çıkarmaya, “doğru” durum dizisinin bulunmaya çalışıldığı bir problemdir. Dejenere modeller için bulunabilecek hiçbir “doğru” durum dizisinin olmadığı açıktır. Bundan dolayı pratik durumlarda, bu problemi en iyi biçimde çözebilmek için genellikle bir optimalite kriterinden faydalanılır. Çok sayıda uygun optimalite kriteri olması nedeniyle, kriterin seçimi, bilinmeyen durum dizisinin belirlenmesinde çok önemli rol oynamaktadır. Optimal durum dizilerinin bulunması ya da her bir durumun ortalama istatistiklerinin elde edilmesi gibi işlemler için kullanımların, modelin yapısından öğrenilmesi gerekebilecektir.

3. problem, belli bir gözlem dizisinin nasıl meydana geldiğini en iyi biçimde tanımlayabilmek için model parametrelerinin optimize edilmeye çalışıldığı bir problemdir. Model parametrelerini ayarlamakta kullanılan gözlem dizisi, saklı Markov modelini “eğitmek” için kullanıldığından eğitim dizisi olarak adlandırılır. Eğitim problemi, model parametrelerinin, gözlemlenen eğitim verilerine optimal olarak ayarlanmasına ve bu sayede en iyi modelin yaratılmasına olanak tanımakta olduğundan saklı Markov modelleri ile ilgili bir çok uygulama için önem taşımaktadır.

Problemlerin çözümleri için geliştirilen yöntemler sırasıyla incelenecektir.

5.1. Saklı Markov Modellerinin Üç Ana Probleminin Çözümü

5.1.1. 1. Problemin çözümü

Bu problemin çözümünde amaç, λ modeli ve $O=O_1, O_2, \dots, O_T$ gözlem dizi varlığında bu gözlem dizisinin elde edilebilme olasılığının hesaplanabilmesidir. Bunu yapmanın en

kestirme yolu, T (gözlem sayısı) uzunluğundaki olası tüm durum dizilerini numaralandırmaktır. q_1 ilk durumu göstermek üzere durum dizileri aşağıdaki şekilde ifade edilebilir.

$$Q = q_1 q_2 \dots q_T \quad (5.1)$$

Eşitlik (5.1)'deki durum dizisi için, gözlemlerin istatistiksel açıdan birbirinden bağımsız olduğu varsayılarak O gözlem dizisinin olasılığı aşağıdaki eşitlikler ile bulunur.

$$P(O|Q, \lambda) = \prod_{t=1}^T P(O_t | q_t, \lambda) \quad (5.2)$$

$$P(O|Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \dots b_{q_T}(O_T) \quad (5.3)$$

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \dots a_{q_{T-1} q_T} \quad (5.4)$$

O gözlem dizisi ve Q durum dizisinin aynı anda olması olasılığı, O ve Q'nun birleşik olasılığı olarak ifade edilir ve Eşitlik (5.3) ile Eşitlik (5.4)'ün çarpımı ile elde edilir.

$$P(O, Q|\lambda) = P(O|Q, \lambda) \cdot P(Q|\lambda) \quad (5.5)$$

O durumunun olasılığı, bu birleşik olasılığın, olası bütün q durum dizileri üzerinden toplamı alınarak elde edilir.

$$P(O|\lambda) = \sum_{\text{tüm } Q} P(O|Q, \lambda) \cdot P(Q|\lambda) \quad (5.6)$$

$$= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \dots a_{q_{T-1} q_T} b_{q_T}(O_T) \quad (5.7)$$

Yukarıda verilen eşitlikteki hesaplamaların yorumu şu şekildedir. Sistem, başlangıç aşamasında (t=1 anında) π_{q_1} olasılığına sahip q_1 durumunda bulunmakta ve $b_{q_1}(O_1)$ olasılığı ile O_1 sinyali üretmektedir. t anından t+1 anına (t=2) geçildiğinde, $a_{q_1q_2}$ olasılığı ile q_1 durumundan q_2 durumuna bir geçiş yaparak $b_{q_2}(O_2)$ olasılığı ile O_2 sembolü üretilir. Bu süreç, $a_{q_{T-1}q_T}$ olasılığı ile q_{T-1} durumundan q_T durumuna geçişi (T anında) yaparak O_T sembolünü üretinceye dek bu şekilde devam eder.

Bu durumda, her bir $t=1,2,\dots,T$ anında ulaşılabilecek N olası durum olması N^T olası durum dizisi olması anlamına gelir ve bu durum dizilerinin her biri için Eşitlik (5.7)' de verilen toplamdaki her bir terim için yaklaşık 2T kadar hesaplamaların gerekli olması nedeniyle $P(O|\lambda)$ 'nin doğrudan doğruya tanımına göre hesaplanabilmesi için $2T \cdot N^T$ dereceden bir hesaplama yapılması gerekecektir. Tam olarak $(2T-1) \cdot N^T$ çarpma ve N^T-1 toplama işlemine ihtiyaç olacaktır. Bu hesaplama N ve T'nin küçük değerleri için bile kestirimsel açıdan elverişli değildir. Örneğin; N=5 (durum sayısı), T=100 (gözlem sayısı) için, $2 \cdot 100 \cdot 5^{100} \approx 10^{72}$ dereceden kestirim vardır. Saklı Markov modellerinin 1. problemi olan Değerlendirme Problemini çözmek için daha etkili bir yönteme gereksinim duyulmuştur. Geliştirilen bu yönteme literatürde ileri-geri yöntemi adı verilmiştir ve doğrudan hesaplama göre daha az karmaşıklıktadır. Bu yöntem ileri ve geri olmak üzere iki şekilde ele alınabilir.

İleri Yöntemi:

λ modeli için $\alpha_i(i)$ ileri değişkeni, t anındaki durumun S_t olduğunda O_1, O_2, \dots, O_t 'nin (t anına kadar) kısmi gözlem dizisi olasılığı olarak tanımlanır.

$$\alpha_i(i) = P(O_1 O_2 \dots O_t, q_t = S_t | \lambda) \quad (5.8)$$

$\alpha_i(i)$ için tümevarım yoluyla aşağıda gösterildiği gibi çözüme ulaşabilir.

1) Başlatma:

$$\alpha_1(i) = P(O_1, q_1 = S_i | \lambda)$$

$$\alpha_1(i) = P(O_1 | q_1 = S_i, \lambda) P(q_1 = S_i)$$

$$\alpha_1(i) = \pi_i b_i(O_1) \quad , \quad 1 \leq i \leq N \quad t=1 \quad (5.9)$$

2) Tümevarım:

$$\alpha_{t+1}(j) = P(O_1 \dots O_{t+1}, q_{t+1} = S_j | \lambda)$$

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \quad , \quad \begin{array}{l} 1 \leq t \leq T-1 \\ 1 \leq j \leq N \end{array} \quad (5.10)$$

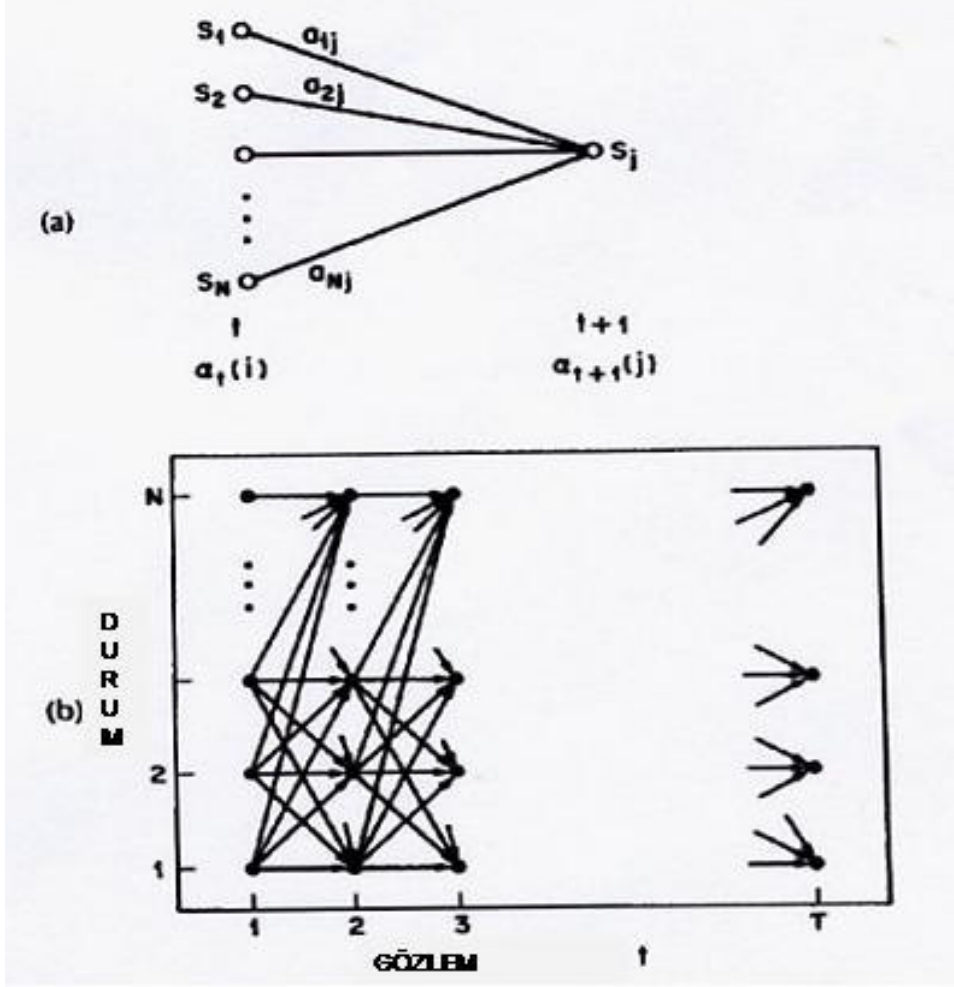
3) Sonlandırma:

$$P(O | \lambda) = \sum_{i=1}^N \alpha_T(i) \quad (5.11)$$

1. adım S_i durumunun ve O_1 ilk gözleminin birleşik olasılığı olarak ileri olasılıkları başlatır.

Eşitlik (5.10)' da $\alpha_t(i)$ ilk gözlemi ve sistemin S_i durumunda olmasını, a_{ij} S_i durumundan S_j durumuna geçiş olasılığını, $b_j(O_{t+1})$ ise $t+1$ anında O_{t+1} gözleminin elde edilmesi olasılığını göstermektedir. Her t anı için tüm durumların ileri değişkenleri hesaplandıktan sonra Eşitlik (5.11) ile gözlemin model tarafından üretilme olasılığı hesaplanır.

İleri hesaplamaların kalbi olan tümevarım adımı, Şekil 5(a)'da gösterilmiştir.



Şekil 5. (a) $\alpha_t(i)$ ileri değişkeninin hesaplanabilmesi için gerekli işlemler dizisinin gösterilişi

(b) t gözlemler ve i durumlar kafesi türünden $\alpha_t(i)$ nin hesaplanması

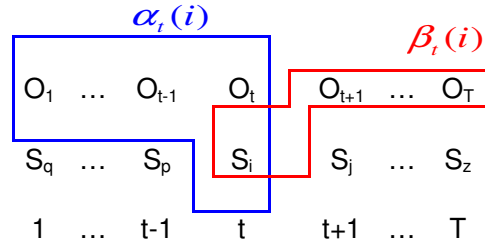
Şekil 4'te, S_i durumuna, t anındaki S_i , $1 \leq i \leq N$ olası durumundan t+1 anında nasıl erişilebileceği gösterilmektedir. $\alpha_t(i)$, O_1, O_2, \dots, O_t ' nin gözlemlenmesi bileşik olayının olasılığı ve t anındaki durum S_i olduğundan, buna göre $\alpha_t(i) a_{ij}$ çarpımı, O_1, O_2, \dots, O_t ' nin gözlemlenmesi bileşik olayının olasılığı olur ve S_j inci duruma, t anındaki S_i durumu vasıtasıyla t+1 anında ulaşılır. $1 \leq i \leq N$ olmak üzere N sayıda olası bütün S_i durumları üzerinde t anında bu çarpımın toplamı alınarak, eşlik eden bütün kısmi gözlemlerle birlikte t+1 anındaki S_j olasılığını vermektedir. Bu yapıldıktan ve S_j biliniyor olduktan sonra j durumundaki O_{t+1} için çözüm yapılarak; yani toplamı alınan büyüklük $b_j(O_{t+1})$ ile çarpılarak

$\alpha_t(i)$ in elde edilebilmektedir. Eşitlik (5.10)' un hesaplanması belli bir t için $1 \leq i \leq N$ olmak üzere bütün j durumları için yapılarak bu hesaplama $t=1,2,\dots,t-1$ için tekrarlanacaktır. Son olarak, 3. Adım, elde edilmek istenilen $P(O|\lambda)$ hesabını, $\alpha_t(i)$ ileri değişkenlerinin toplamı olarak verecektir.

$$\alpha_T(i) = P(O_1 O_2 \dots O_T, q_T = S_i | \lambda) \quad (5.12)$$

$1 \leq t \leq T$ ve $1 \leq j \leq N$ olmak üzere, $\alpha_T(j)$ 'nin hesaplanması kullanılarak doğrudan hesaplama yönteminde ihtiyaç duyulan $2TN^T$ derece yerine N^2T dereceden hesaplamalar gerektirdiği görülür. Tam olarak, $N(N+1)(T-1)+N$ sayıda çarpıma ve $N(N-1)(T-1)$ sayıda toplama işlemi yapılmaktadır. Örneğin, $N=5$ ve $T=100$ için ileri yöntemine göre 3000 hesaplama ve doğrudan hesap yöntemine göre ise 10^{72} adet hesaplama yapılması gereklidir. Bu durumda derece büyüklüğüne göre ileri yöntemi kullanılarak yaklaşık 69 kat tasarruf edilmiş olur.

İleri olasılık hesabı, esasen Şekil 4(b)' de gösterilmiş olan kafes yapısı üzerine kuruludur. Buradaki ana düşünce, gözlem dizisi ne kadar uzun olursa olsun, düğümler her defasında kafes içindeki aralıklarda olmak üzere sadece N sayıda durum olduğundan, olası bütün durum dizilerinin bu N düğüm içinde yeniden şekillenecek olmasıdır. $t=1$ anında, (kafesteki ilk zaman aralığı) $1 \leq i \leq N$ olmak üzere $\alpha_T(i)$ değerlerinin hesaplanması gerekmektedir. $t=2,3,\dots,T$ olduğunda ise, her bir hesaplamada sadece bir önceki zaman aralığındaki N sayıda $\alpha_{T-1}(i)$ değeri gerekli olacağından, N sayıdaki her bir grid noktaya, bir önceki zaman aralığındaki N noktadan ulaşıldığından, $1 \leq i \leq N$ olmak üzere sadece $\alpha_T(j)$ değerlerinin hesaplanması yeterli olacaktır.



Şekil 6. İleri ve geri değişkenlerinin gösterilişi

Geri Yöntemi:

Geri yöntemi için Şekil 5'te gösterildiği gibi, verili bir λ modelinde t zamanındaki durum S_i olmak üzere, $t + 1$ 'den itibaren son ana kadar kısmi gözlem dizilerinin olasılığını ifade edecek biçimde aşağıdaki biçimde tanımlanan $\beta_t(i)$ geri değişkeni düşünülebilir.

$$\beta_t(i) = P(O_{t+1}O_{t+2}\dots O_T | q_t = S_i, \lambda) \quad (5.13)$$

Burada aşağıdaki gibi, tümevarım yöntemiyle $\beta_t(i)$ için çözüm yapılabilir.

1) Başlatma:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (5.14)$$

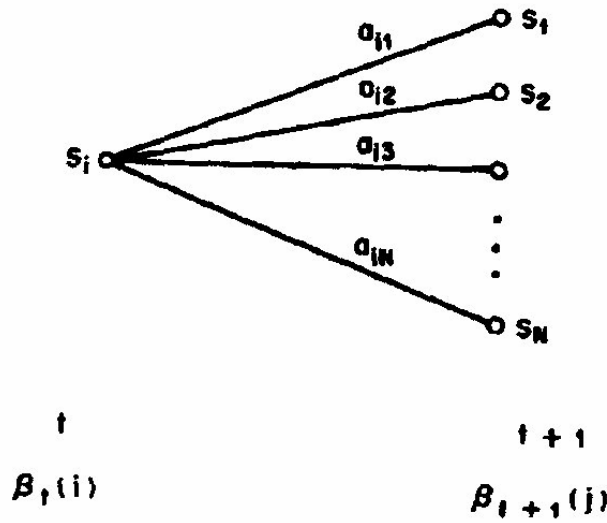
2) Tümevarım:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t=T-1, T-2, \dots, 1, \quad 1 \leq i \leq N \quad (5.15)$$

3) Sonlandırma:

$$P(O|\lambda) = \sum_{i=1}^N \pi_i \beta_1(i)$$

Başlangıçtaki 1. Adım'da, bütün i 'ler için, $\beta_T(i)$ 'ler rasgele biçimde 1 olarak tanımlanır. Şekil 5' te çizimle açıklanmış olan 2. Adım, t anında S_i durumunun içinde olunabilmesi ve $t+1$ anından itibaren gözlem dizisinin açıklanabilmesi için, j durumundaki O_{t+1} gözlemi ($b_j O_{t+1}$ terimi) ile birlikte S_i 'den S_j 'ye geçişini de (a_{ij} terimi) hesaba katıp, daha sonra da j durumundan arta kalan kısmi gözlem dizisini ($\beta_{T+1}(j)$ terimi) hesaba katarak, $t+1$ anındaki olası bütün S_j durumlarının göz önünde bulundurulması gerekmektedir.



Şekil 7. $\beta_T(i)$ geri değişkeninin hesaplanabilmesi için gerekli işlemler dizisinin gösterilişi

2. ve 3. problemlerin çözümünde de kullanılan geri değişken $\beta_T(i)$ ' nin hesaplanması için, $1 \leq t \leq T$ ve $1 \leq i \leq N$ olmak üzere N^2T dereceden hesaplamalar gerekli olup bu işlem Şekil 4 (b)'dekine benzer bir kafes yapısı yardımıyla yapılabilmektedir.

5.1.2. 2. Problemin Çözümü

Ekstra bir çözümün verilebileceği 1. Problemden farklı olarak, 2. Problemi çözenin, yani belirli bir gözlem dizisine ait "optimal" durum dizisini tespit etmenin birkaç yolu vardır. Bu problemde karşılaşılan en büyük zorluk, optimal durum dizisini tanımlanmasıdır. Birden çok olası optimallik kriteri olmasıdır. Örneğin, optimallik kriterlerinden biri, tek başlarına en çok olası olan q_t durumlarını seçmektir. Bu optimallik kriteri beklenen her bir doğru durum sayısını maksimize edecektir. Bu çözümü 2. Problemin çözümüne uygulamak için, bir λ modeli ve O_t gözlem dizisi için t anında S_i durumu içinde bulunması olasılığı aşağıda Eşitlik (5.16)'da verildiği biçimde tanımlanır.

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) \quad (5.16)$$

$\alpha_t(i)$ değişkeni, O_1, O_2, \dots, O_t kısmi gözlem dizisini ve t anındaki S_i durumunu hesaba kattığından ve $\beta_T(j)$ de t anında verili bir S_j durumu için $O_{t+1}, O_{t+2}, \dots, O_T$ kısmi gözlem dizisinin geri kalan kısmını hesaba kattığından, Eşitlik (5.16)'daki eşitlik basitçe ileri geri değişkenler biçiminde Eşitlik (5.17)'deki gibi ifade edilebilir.

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_t(i)\beta_t(i)} \quad (5.17)$$

$P(O|\lambda) = \sum_{i=1}^N \alpha_t(i)\beta_t(i)$, $\alpha_t(i)$, $\beta_t(i)$ normalizasyon faktörü, $\gamma_t(i)$ 'yi aşağıdaki biçimde tanımlanacak bir olasılık ölçüsü haline getirir.

$$\sum_{i=1}^N \gamma_t(i) = 1 \quad (5.18)$$

Eşitlik (5.17)'de $\alpha_t(i)$, $\beta_t(i)$ çarpımı tüm gözlem dizisini 2 parçaya bölüp tamamını taramaktadır ve bu çarpım normalize edilerek her t anında Eşitlik (5.18)'in gerçekleşmesi sağlanır.

$\gamma_t(i)$ 'den faydalanarak, t anındaki tek tek alındığı haldeki en olası durum q_t için aşağıdaki eşitlik ile çözüme ulaşılır.

$$q_t = \arg \max_{1 \leq i \leq N} [\gamma_t(i)] \quad , \quad 1 \leq t \leq T \quad (5.19)$$

Eşitlik (5.19)'da her ne kadar, beklenen doğru durum sayısını (her bir t anı için en olası durumu seçerek) maksimize edilse de, sonuç olarak bulunan durum dizisi ile ilgili bazı problemler olacaktır. Örneğin, saklı Markov modellerinde bazı i ve j değerleri için $a_{ij}=0$ olduğunda sıfır olasılığa sahip durum geçişlerinin olması durumunda "optimal" durum dizisi, gerçekte, geçerli bir durum dizisi bile olmayabilir. Bunun nedeni, Eşitlik (5.19)'daki çözümünün, durum dizilerinin oluşumunu dikkate almaksızın, sadece her bir andaki en olası durumu belirlemesidir.

2. Problemin olası bir çözümü, optimallik kriterini modifiye etmektir. Örneğin, beklenen doğru durum çiftlerinin (q_t, q_{t+1}) ya da durum üçlülerinin (q_t, q_{t+1}, q_{t+2}) vb. sayısı maksimize edilecek biçimde çözüm yapılabilir. Bu kriterler bazı uygulamalar için geçerli olabilse de, en çok kullanılan kriter tekil en iyi durum dizisini tespit etmek; yani, $P(Q, O | \lambda)$, 'yı maksimize etmekle eşdeğer olan $P(Q | O, \lambda)$ 'yı maksimize etmektir. Bu iyi tekil durum dizisini bulmak için, dinamik programlama metoduna dayalı biçimsel bir teknik bulunmakta olup, Viterbi algoritması olarak adlandırılmaktadır (Rabiner, 1989).

Viterbi Algoritması :

Verilmiş bir $O = \{O_1, O_2, \dots, O_T\}$ gözlem dizisi için en iyi tekil durum dizisi $Q = \{q_1, q_2, \dots, q_t\}$ 'yi bulmak için ilk t gözlemlerini ifade etmekte olan ve S_i içinde sonlanan, t anında tek bir yol üzerindeki en büyük skor (en yüksek olasılığı) olarak $\delta_t(i)$ 'nin aşağıdaki biçimde tanımlanması gereklidir.

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1 q_2 \dots q_t = i, O_1 O_2 \dots O_t | \lambda] \quad (5.20)$$

Tümevarım yoluyla Eşitlik (5.21) elde edilir

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) a_{ij} \right] \cdot b_j(O_{t+1}) \quad (5.21)$$

Durum dizisini tekrar eski haline getirmek için, her t ve j için (5.21)'deki ifadeyi maksimize hale getirmiş olan argümanın sürekli olarak izlenmesi gerekir. Bu, $\psi_t(j)$ sıralı dizisini kullanarak yapılır. Bu durumda en iyi durum dizisini bulmak için gerekli yöntem şu şekilde ifade edilebilir :

1) Başlatma:

$$\delta_1(i) = \pi_i b_i(O_1) , \quad 1 \leq i \leq N \quad (5.22a)$$

$$\psi_1(i) = 0 \quad (5.22b)$$

2) Yineleme:

$$\delta_t(j) = \max_{1 \leq i \leq N} \left[\delta_{t-1}(i) a_{ij} \right] b_j(O_t) , \quad 2 \leq t \leq T$$

$$1 \leq j \leq N \quad (5.23a)$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} \left[\delta_{t-1}(i) a_{ij} \right] , \quad 2 \leq t \leq T$$

$$1 \leq j \leq N \quad (5.23b)$$

3) Sonlandırma

$$P^* = \max_{1 \leq i \leq N} \left[\delta_T(i) \right] \quad (5.24a)$$

$$q_T^* = \arg \max_{1 \leq i \leq N} \left[\delta_T(i) \right] \quad (5.24b)$$

4) Yol (durum dizisi) geri izleme :

$$q_t^* = \psi_{t+1}(q_{t+1}^*) , \quad t = T-1, T-2, \dots, 1 \quad (5.25)$$

Viterbi algoritmasının, uygulamada İleri Yöntemindeki Eşitlik (5.9) - (5.11) hesaplamalarına (geri izleme adımı hariç) benzer olduğuna dikkat edilmelidir. En büyük farklılık, (5.10)'daki toplama yöntemi yerine kullanılan (5.23a)'daki maksimizasyondur. t anında j. durum için hesaplanıyorken bu durumun öncesinde olabilecek tüm i. durumlar ele alınmak yerine sadece en büyük olasılığa sahip olan durum hesaba katılır(Haberdar, 2005). Kafes ya da kamerye şeklindeki bir yapı Viterbi yönteminin hesaplanışını etkili bir biçimde yürütmektir.

5.1.3. 3. Problemin Çözümü

Saklı Markov modelleriyle ilgili problemlerin üçüncüsü ve en zoru olan 3. Problem, model parametrelerini verilen bir modele göre gözlem dizisinin olasılığını maksimize edecek biçimde ayarlamak için bir metot belirlemektir. Gözlem dizisi olasılığını maksimize edecek model için analitik olarak çözüm yapmanın bilinen bir yolu mevcut değildir. Aslında, eğitim verileri olarak verilen sonlu sayıdaki gözlem dizisi için, model parametrelerini kestirmenin optimal bir yolu yoktur. Yine de, Baum - Welch metodu gibi yinelemeli bir metot (ya da eşdeğer bir EM (tahmin - modifikasyon) metodu veya gradyan teknikleri kullanarak, $P(O|\lambda)$ lokal olarak maksimize olacak biçimde $\lambda=(A,B,\pi)$ seçimi yapabilir. Tez çalışmasının bu bölümünde, model parametrelerinin seçilmesi ile ilgili olarak, temelde Baum ve arkadaşlarının klasik çalışmalarına dayanarak bir yineleme yöntemi ele alınmıştır.

Saklı Markov modeli parametrelerinin yeniden kestirimi yinelemeli güncelleme ve düzeltme yöntemini tanımlayabilmek amacıyla öncelikle, verili bir gözlem ve durum dizisi için, t anında S_i durumunda ve t+1 anında da S_j durumunda olma olasılığı aşağıdaki

biçimde tanımlanan $\xi_t(i, j)$ değişkeninin tanımlanması gerekmektedir.

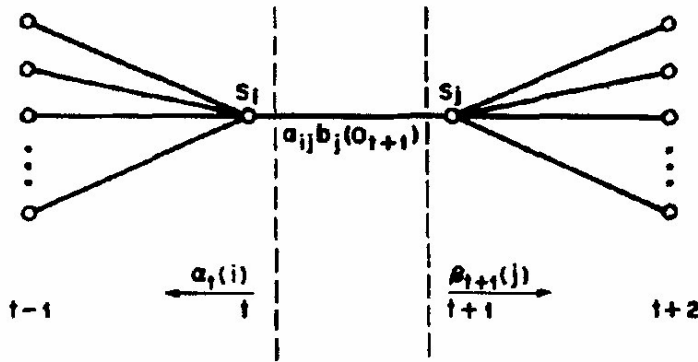
$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \quad (5.26)$$

Eşitlik (5.26)'ya göre gerekli olan koşullara yol açan olaylar dizisi Şekil 6' da gösterilmiştir.

İleri ve geri, değişkenlerin tanımından açıkça anlaşılacağı gibi, $\xi_t(i, j)$, yi, paydaki terim

sadece $P(q_t = S_i, q_{t+1} = S_j | O, \lambda)$ olacak ve $P(O | \lambda)$ ile bölüm yapıldıktan sonra istenilen olasılık önlemi elde edilecek biçimde Eşitlik (5.27)' de verildiği şekilde düzenlenir.

$$\begin{aligned} \xi_t(i, j) &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O | \lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \end{aligned} \quad (5.27)$$



Şekil 8. Sistemin t anında S_i ve t+1 anında ise S_j içinde bulunması bileşik olasılığının hesaplanması için gerekli işlemler dizisinin gösterilmesi

$\gamma_t(i)$, verili bir gözlem dizisi ve model için t anında S_i durumunda bulunulması olasılığı olarak tanımlanmış olduğu için, $\gamma_t(i)$ ile $\xi_t(i, j)$ arasında,

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (5.28)$$

biçimde bir ilişki kurulabilir. t zaman indisi üzerinde $\gamma_t(i)$ için toplama yapıldığında (t = T zaman aralığı toplama dahil edilmediğinde), S_i durumunun beklenen ziyaret edilme sayısı ya da bununla eşdeğer anlamda S_i durumundan yapılması beklenen geçişlerin sayısı olarak da yorumlanabilecek bir nicelik elde edilir. Benzer biçimde,

$\xi_t(i, j)$ 'nin t (t=1'den t=T-1'e kadar) üzerinde toplamı, S_i durumundan S_j durumuna beklenen geçiş sayısı olarak yorumlanabilecektir.

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{Beklenen } S_i \text{'den geçiş sayısı} \quad (5.29a)$$

$$\sum_{t=1}^{T-1} \xi_t(i, j) = \text{Beklenen } S_i \text{'den } S_j \text{'ye geçiş sayısı} \quad (5.29b)$$

Yukarıdaki formüller ve olayların oluşumu kullanılarak, herhangi bir saklı Markov modeline ait parametrelerin yeniden kestirimi için bir metodun tanımı verilebilir. $\bar{\pi}$, A ve B için geçerli yeniden kestirim formülleri kümesi aşağıdaki biçimde ifade edilmektedir.

$$\bar{\pi}_i = (t=1) \text{ anında } S_i \text{ durumunda beklenen sıklık} = \gamma_1(i) \quad (5.30a)$$

$$\bar{a}_{ij} = \frac{\text{Beklenen } S_i \text{'den } S_j \text{'ye geçişlerin sayısı}}{\text{S}_i \text{'den geçişlerin beklenen sayısı}}$$

$$= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (5.30b)$$

$\bar{b}_j(k) =$ Beklenen j durumundaki ve v_k sembolünün gözlenme sayısı
j durumunda beklenen olay sayısı

$$\bar{b}_j(k) = \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(i)} \quad (5.30c)$$

Baum-Welch algoritmasının adımları aşağıda verilmiştir:

- Mevcut $\lambda = (A, B, \pi)$ model parametreleri kullanılarak $\xi_t(i, j)$ ve $\gamma_t(i)$ değerleri hesaplanır.
- Belirlenen $\xi_t(i, j)$ ve $\gamma_t(i)$ değerleri ile yeni parametre seti $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$ elde edilir.
- $P(O|\bar{\lambda}) > P(O|\lambda)$ olduğu sürece ilk adıma geri dönülür.

Adımlar uygulandığında mevcut model $\lambda = (A, B, \pi)$ biçiminde tanımlanır ve (5.30a) - (5.30c) eşitliklerinin sağ tarafını hesaplamak için kullanılır ve de yeniden hesaplanmış model (5.30a)-(5.30c) eşitliklerinin sol tarafına göre belirlendiği şekliyle $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$ olarak tanımlanırsa, Baum ve arkadaşları tarafından kanıtlanmış olan $\bar{\lambda} = \lambda$ ilk model olacak biçimde olabilirlik fonksiyonunun kritik bir noktası tanımlanacak, ya da $\bar{\lambda}$ modeli, $P(O|\bar{\lambda}) > P(O|\lambda)$ olması açısından λ modelinden daha çok olabilir olacaktır. Bu yöntem kullanılarak gözlem dizisinin üretilmesinin daha çok olası olduğu yeni bir λ modeli bulmuş olunur.

Yukarıdaki yönteme dayanarak, λ yerine $\bar{\lambda}$ yinelemeli olarak kullanarak kestirme hesabı tekrarlandığında, belli bir sınır noktasına ulaşıncaya dek O gözlemler dizisinin modelden gözlemlenebilmesi olasılığı iyileştirilebilir. Bu yeniden kestirim yönteminin sonucuna, "Saklı Markov modelinin maksimum olabilirlik kestirimi" adı verilir. Burada, İleri - geri

algoritmasının sadece lokal maksimuma götüreceğine ve birçok önemli problemde optimizasyon yüzeyinin son derece kompleks olduğuna ve birçok lokal maksimuma sahip olduğuna dikkat edilmelidir.

Eşitlik (5.30a) ve Eşitlik (5.30c) yeniden kestirim formülleri, standart zorunlu optimizasyon teknikleri kullanılarak Baum'un yardımcı fonksiyonu $\bar{\lambda}$ üzerinde maksimize edilerek Eşitlik (5.31) ile doğrudan doğruya türetilebilecektir.

$$Q(\lambda, \bar{\lambda}) = \sum_Q P(Q|O, \lambda) \log [P(O, Q|\bar{\lambda})] \quad (5.31)$$

Baum tarafından $Q(\lambda, \bar{\lambda})$ 'nin optimizasyonunun daha yüksek bir olabilirlik sonucunu verdiği gösterilmiştir.

$$\max_{\bar{\lambda}} [Q(\lambda, \bar{\lambda})] \Rightarrow P(O|\bar{\lambda}) \geq P(O|\lambda) \quad (5.32)$$

Sonuç olarak olabilirlik fonksiyonu kritik bir noktaya doğru yakınsar.

6. SONUÇLAR

Tez çalışmasında özellikle konuşma sinyallerinde son yıllarda yoğun olarak kullanılan saklı Markov modelleri ele alınmıştır. Saklı Markov modellerinin yapısal güçlüğü ve matematiksel alt yapısının karmaşıklığı model parametrelerine ilişkin tahminleri zorlaştırmaktadır. Bu nedenle bu modeller güçlü matematiksel altyapı ve bilgisayar donanımı gerektirmektedir.

Tez çalışması, saklı Markov modelleri için sadece giriş düzeyinde bir çalışma olarak değerlendirilmelidir. Burada amaç, saklı Markov modellerinin bir istatistikçi gözü ile incelenmesidir. Bu nedenle tez çalışmasında herhangi bir model denemesi yetersiz donanım ve veri eksikliği nedeniyle yapılamamıştır. Burada, saklı Markov modellerinde gözlenen üç problem detaylı olarak ele alınmıştır.

KAYNAKLAR

- Allison L., Hidden Markov Models, School of Computer Science and Software Engineering, Monash University, Australia
- Bhat, U. N., 1972, Elements of Applied Stochastic Processes, Wiley
- Boyle, R. D., Hidden Markov Models,
http://www.comp.leeds.ac.uk/roger/HiddenMarkovModels/html_dev/main.htm
- C. D. Manning, H. Schütze. Foundations of Statistical Natural Language Processing. MIT Press, 1999, 680 p.
- Elliot, R. J., Aggoun L., Moore J.B., 1995, Hidden Markov Models: Estimation and Control, Springer- Verlag, New York
- Froetschl B., 1998, Hidden Markov Models (HMMs), <http://www-agki.tzi.de/ik98/prog/kursunterlagen/t2/node4.html#SECTION0002200>
- Güvenç, G., 1995, Saklı Markov Modelleri ile Konuşmacıdan Bağımsız Ayrık Sözcükleri Tanıma, Y. Lisans Tez, Elektrik-Elektronik Mühendisliği Bölümü, Hacettepe Üniversitesi
- Haberdar, H., 2005, Saklı Markov Model Kullanılarak Görüntüden Gerçek Zamanlı Türk İşaret Dili Tanıma Sistemi, Y. Lisans Tez, Bilgisayar Mühendisliği Bölümü, Yıldız Teknik Üniversitesi
- Hidden Markov Model, www.free-definition.com
- Huang X. D. ve Ariki Y., Hidden Markov Models for Speech Recognition. Edinburgh University Press, 1990, 275 p.
- İnal C., 1988. Olasılıksal Süreçlere Giriş, Hacettepe Üniversitesi Yayınları A-56, Ankara, 1-11s.
- Rabiner L. ve Juang B.H., Fundamentals of Speech Recognition. Prentice Hall, 1995, 507 p.
- Moscow State University, Dept. of Computational Mathematics & Cybernetics, Hidden Markov Models, <http://leader.cs.msu.su/~luk/HMM.html>
- Rabiner, L.R., 1989, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, Vol. 77, No.2
- Rabiner L. ve Juang B.H., Fundamentals of Speech Recognition. Prentice Hall, 1995, 507 p.
- Thyer M., Kuczera G., 2003, A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach, Journal of Hydrology 275, 12-16
- Vaseghi, S. V., 2000, Advanced Digital Signal Processing and Noise Reduction, Second Edition

EK

1. Problemin çözümü için geliştirilmiş Turbo Pascal algoritması aşağıda sunulmuştur.

```
uses crt;
var
  A:array[1..100,1..100] of integer;
  B:array[1..100,1..100] of integer;
  Aft:array[1..100] of real;
  Aftl:array[1..100] of real;
  P: array[1..100] of real;
  top,i,j:integer;
  Pox:longinteger;
  M,N,T,Z:integer;
begin
  clrscr;
  write('Durum sayısını giriniz:');readln(N);
  write('Durumlardaki sembol sayısını giriniz:');readln(M);
  write('Gözlem sayısını giriniz:');readln(T);
  for i:=1 to N do
    begin
      write(i,'. durum başlangıç değerini girin:');readln(P[i]);
    end;
  for i:=1 to N do
    begin
      for j:=1 to M do
        begin
          write(i,'. durum',j,'. Sembol olasılığını girin:');readln(B[i,j]);
        end;
      end;
    end;
  for i:=1 to N do
    begin Aft[i]:=P[i]*B[i,1];
    end;
  top:=0;
  for i:=1 to M do
    begin
      for i:=1 to N do
        begin
          top:=Aft[j]*A[i,j]*B[j];
          Aftl[j]:=Aftl[j]+top;
        end;
      end;
    end;
  Pox:=0;
  for i:=1 to Ai do
    begin
      Pox:=Aft[i]+Pox;
    end;
  writeln('Sonuç:Pox);
  readln;
end.
```

ÖZGEÇMİŞ

Adı Soyadı : Züleyha SEMERCİ
Doğum Yeri : Ankara
Doğum Yılı : 1981
Medeni Hali : Bekar

Eğitim ve Akademik Durumu

Lise : 1998 Mustafa Kemal Lisesi- Yenimahalle ANKARA
Lisans : 1998-2005 HÜ Fen Fakültesi İstatistik Bölümü
Yabancı Dil : İngilizce

İş Tecrübesi

2004-2005 : Umut Tanı Sağlık Matbaa A.Ş.
2005- : Oyakbank A.Ş.