

167759

T.C
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

WEB SİTELERİ İÇİN YAPAY BAĞIŞIKLIK TABANLI
BİR ÖNERİ SİSTEMİ

Ayşe Merve ŞAKİROĞLU

YÜKSEK LİSANS TEZİ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Konya, 2005

T.C.
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

WEB SİTELERİ İÇİN YAPAY BAĞIŞIKLIK TABANLI
BİR ÖNERİ SİSTEMİ

Ayşe Merve ŞAKİROĞLU

YÜKSEK LİSANS TEZİ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Bu tez 15/08/2005 tarihinde aşağıdaki jüri tarafından oybirliği ile kabul edilmiştir



Prof. Dr. Ahmet ARSLAN
(A.B.D Bşk.- Danışman)



Prof. Dr. Ferruh YILDIZ
(Üye)



Doç. Dr. Şirzat KAHRAMANLI
(Üye)

ÖZET

Yüksek Lisans Tezi

WEB SİTELERİ İÇİN YAPAY BAĞIŞIKLIK TABANLI BİR ÖNERİ SİSTEMİ

Ayşe Merve ŞAKIROĞLU

Selçuk Üniversitesi Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof.Dr.Ahmet ARSLAN

2005, 74 Sayfa

Jüri : Doç.Dr.Şirzat KAHRAMANLI

Prof.Dr.Ahmet ARSLAN

Prof.Dr. Ferruh YILDIZ

Öneri sistemleri (Recommender Systems), alışveriş ve eğlence sitelerinde kullanıcının yolunu bulması için tasarlanmış Internet tabanlı yazılım araçlarının yeni bir çeşididir. Öneri sistemlerinin amacı o an sitede aktif durumda bulunan kullanıcıların profiline benzer profilleri bulmak ve bu doğrultuda kullanıcıya tavsiye verip yönlendirmektir.

Bu tezde yapay bağışıklık (Artificial Immune System) tabanlı bir öneri sistemi tasarlanmıştır. Sistem tasarlanırken ilk olarak web sitesini ziyaret eden

kullanıcıların profilleri belirlenmiş ve benzer profiller bulunup komşuluklar oluşturulmuştur. Bunun için bir yapay bağıklık algoritması olan AINET tabanlı bir model kurulmuştur. AINET'in kullanılmasının sebebi öneri sistemlerinin başlıca sorunları olan bilgi eksikliği ve ölçeklenebilirlik problemlerinin her ikisinin de giderilebilecek potansiyele sahip olmasındandır.

Tez için geliştirilen uygulamada sunulan model diğer yöntemlerle kıyaslanmıştır. Değerlendirme yöntemi olarak Mutlak Hata Ortalaması ve Çağrılan Ölçütü kullanılmıştır. Uygulamanın sonucunda model için uygun parametre değerleri seçildiğinde modelin diğer yöntemler ile benzer hata seviyesinde tahmin ürettiği gözlemlenmiştir. Buna karşın yapılan tavsiyenin doğruluğunda ve hızında sağladığı artış sunulan modelin kullanımının daha avantajlı olduğunu göstermiştir.

Anahtar Kelimeler: Öneri Sistemleri, Yapay Bağıklık Sistemi, aiNET, İşbirlikçi Filtreleme Tekniği

ABSTRACT

Master Thesis

A RECOMMENDER SYSTEM BASED ON ARTİFİCİAL IMMUNE FOR WEB SİTES

Ayşe Merve ŞAKİROĞLU

Selçuk University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering problem

Supervisor: Prof.Dr.Ahmet ARSLAN

2005, 74 Page

Jury: Assoc.Prof.Dr.Şirzat KAHRAMANLI

Prof.Dr.Ahmet ARSLAN

Prof.Dr. Ferruh YILDIZ

Recommender systems are new type of internet based software tools. These are designed for helping users to find their path on the shopping and entertainment web sites. Their purpose are finding similar profiles with the active user's profile that is online on the site at that time and giving recommendations according to these similarities.

In this thesis, a recommendation system based on artificial Immune systems is designed. Firstly the users' profiles have been learned then similarities between profiles detected by using history votes given by users to items and generated neighbourhoods. For this, a model has been constructed based on an AIS algorithm, AINET which has got a potential for solving mainly trouble of the recommender systems, sparsity and scalability problems.

In the application developed for this thesis, proposed model and other methods are compared. MAE (Mean Absolute Error) and Recall metrics is used as an evaluation metric. It's observed that results of application has provided prediction with an error rate similar to other methods when convenience parameter values have been selected. But proposed model has generated recommendations more accuracy and speedy than the other models. So it shows that the model based on AINET has provided more advantages comparatively others.

Key Words: Recommender System, Artificial Immune System, aiNET, Collaborative Filtering Technique

TEŐEKKÜR

Yüksek lisansım boyunca katkılarını, yönlendirici desteğini ve anlayışını hiçbir zaman esirgemeyen danışmanım Sayın Prof. Dr. Ahmet ARSLAN'a ; çalışmalarım sırasında yardımlarını ve manevi desteğini esirgemeyen Sayın Doç.Dr. Şirzat KAHRAMALI'ya; bu süre boyunca göstermiş oldukları manevi desteklerinden, ilgilerinden ve yardımlarından dolayı Arş.Gör. Semiye YILMAZ'a, Arş.Gör. Humar KAHRAMANLI'ya , Arş.Gör. Süheyla YILDIZ'a ve Bilgisayar Mühendisliği Bölümü'nün tüm öğretim elemanlarına, çalışmalarım boyunca manevi ve maddi destekleri ile her zaman yanımda olan aileme çok teşekkür ederim.



İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	iii
TEŞEKKÜR.....	v
İÇİNDEKİLER	1
ŞEKİLLER ve TABLOLAR.....	3
1. GİRİŞ	5
1.1 Tezin Konusu, Amacı Ve Önemi	7
1.2 Literatür Özeti	9
2. ÖNERİ SİSTEMLERİ	11
2.1 Öneri Sistemlerinin Yapısı.....	12
2.1.1 İçerik Tabanlı Filtreleme Teknikleri	12
2.1.2 İşbirlikçi Filtreleme Teknikleri	13
2.1.2.1 Komşulukların Oluşturulması ve Seçimi	15
2.1.2.2 Tahmin ve Tavsiye Üretme.....	20
2.1.2.3 Değerlendirme Ölçütleri.....	22
2.1.2.3.1 Tahmin Değerlendirme Ölçütleri	22
2.1.2.3.2 Tavsiye Değerlendirme Ölçütleri	26
2.1.2.3.3 Öneri Sisteminin Performansının Değerlendirilmesi	27
2.1.3 Karşılaşılan Problemler ve Sunulan Çözümler	27
2.1.3.1 Bilgi Eksikliği	28
2.1.3.2 Ölçeklenebilirlik.....	29
3. YAPAY BAĞIŞIKLIK SİSTEMLERİ	32
3.1 İnsan Bağışıklık Sistemi.....	32
3.1.1 İnsan Bağışıklık Sisteminin Yapısı	32
3.1.2 İnsan Bağışıklık Sisteminin Mimarisi.....	33
3.1.3 Bağışıklık Hücreleri	35
3.1.3.1 Lenfositler	35
3.1.3.1.1 B hücreleri ve antikorlar.....	35
3.1.3.1.2 T hücreleri	37
3.1.3.2 Fagositler, Granülositler ve onların ilişkileri	37
3.1.4 Bağışıklık Sistemi Vücudu Nasıl Korur ?	38
3.1.5 Klonal Seçme Teorisi.....	39
3.1.6 Bağışıklık Ağ Teorisi	41
3.2 Yapay Bağışıklık Sistemleri.....	43
3.2.1 Şekil Uzayı.....	45
3.2.2 Duyarlılık Hesaplaması	47
3.2.3 Bağışıklık Algoritmaları.....	48
3.2.3.1 Klonal Seçme Algoritması	49
3.2.3.2 Bağışıklık Ağ Modelleri.....	50
3.2.3.2.1 aiNET Algoritması	51
4. YAPAY BAĞIŞIKLIK TABANLI BİR ÖNERİ SİSTEMİ MODELİ	56
4.1 MovieLens Veri Kümesi ve Kullanıcıların Kodlaması.....	56
4.2 Sunulan Öneri Modeli	57

4.3 Uygulama	61
4.3.1 Çalışma Mantığı	61
4.3.2 Uygun Parametre Değerlerinin Tespiti	62
4.3.3 Sunulan Modelin Diğer Yöntemlerle Kıyaslanması	67
5.SONUÇLAR	71
6. KAYNAKLAR	72



ŞEKİLLER ve ÇİZELGELER

Şekil 2.1: Öneri Verme Süreci	11
Şekil 2.2: İşbirlikçi filtreleme süreci (Sarwar B.M. ve ark. 2001).....	14
Şekil 2.3 : Komşuluk Gösterimi. Aktif kullanıcı ile diğer kullanıcılar arasındaki uzaklık gösterilmiş ve en yakın k (örnek için 5) tane kullanıcı komşu olarak belirlenmiştir (Sarwar B.M. ve ark. 2000).	16
Şekil 2.4 : (a): kullanıcıya en yakın n tane kullanıcının komşu olarak seçilmesi.	18
(b): küme merkezine en yakın n tane kullanıcının komşu olarak seçilmesi.	18
Çizelge 2.1 : Kümeleme ile komşu seçimini kullanan İFT algoritması.....	19
Şekil 2.5 : Kümeleme yardımı ile komşu seçimi	20
Çizelge 2.2 : Tavsiye Değerlendirme Ölçütleri.....	27
Şekil 2.6 : Boyut azaltmanın yararını belirtmek için tasarlanmış örnek yapı	30
Şekil 2.7 : Boyut azaltma uygulandıktan sonra tasarlanmış örnek yapının hali	30
Şekil 3.1 : Bağışıklık Sisteminin Sınıflandırılması ve Temel elemanları	32
Şekil 3.2 : Bağışıklık Sisteminin Çok Katmanlı Mimarisi.....	34
Şekil 3.3 Antikor'un Yapısı	36
Şekil 3.4 Bağışıklık Sisteminin basit aktivasyon mekanizması	39
Şekil 3.5 Klonal Seçme prensibi	40
Şekil 3.6 : Bağışıklık Ağ teorisinin basit kavramları. (a) Farklı B hücreleri tarafından tanınan çeşitli epitoplara sahip bir antijendir. (b) Antikorun V bölgesi ile paratop ve idiotopun gösterimidir.	41
Şekil 3.7 : Bağışıklık Ağ teorisinin gösterimi. (a) Bir antijenin 1 nolu sınıfın bir antikorunu uyarması, bu antikorun 2 nolu sınıfın antikorunu uyarması vb. (b) İdiyotipik ağ yapısının ayrıntılı gösterimi.....	42
Şekil 3.8 YBS' nin Katmanlı Yapısı	44
Şekil 3.9: Şekil Uzayı Gösterimi.....	46
Şekil 3.10 : Farklı bir şekil uzayı gösterimi (Tanıma çemberi)	47
Çizelge 3.1 : Duyarlılık Ölçütü için kullanılacak Uzaklıklar.....	48
Şekil 4.1 : aiNet Tabanlı Öneri Sistemi	62
Şekil 4.2 : <i>ölüm_eşiği</i> değerine göre MHO değerinin değişimi.....	63
Şekil 4.3 : <i>ölüm_eşiği</i> değerine göre Çağrılan değerinin değişimi	63

Şekil 4.4 : <i>ağ_baskılama_eşiği</i> değerine göre MHO değerinin değişimi	64
Şekil 4.5 : <i>ağ_baskılama_eşiği</i> değerine göre Çağrılan değerinin değişimi.....	65
Şekil 4.6 : <i>n birey</i> değerine göre MHO değerinin değişimi	65
Şekil 4.7 : <i>n birey</i> değerine göre Çağrılan değerinin değişimi.....	66
Şekil 4.8 : ζ değerine göre MHO değerinin değişimi	67
Şekil 4.9 : ζ değerine göre Çağrılan değerinin değişimi.....	67
Şekil 4.10 aiNet, k-ortalama, standart İFT yöntemlerinin MHO ölçüt değerlerinin k sabitine göre değişimi	68
Şekil 4.11 aiNet ve Standart İFT yönteminin çağrılan ölçüt değerlerinin k sabitine göre değişimi.....	69
Şekil 4.12 aiNet ve Standart İFT yönteminin tavsiye zamanının k sabitine göre değişimi.....	69



1. GİRİŞ

Teknolojinin hızla gelişmesi ve kullanımının ucuzlaması veri depolamayı oldukça kolaylaştırmış hatta işletmeler için verilerin saklanması ihtiyacı haline getirmiştir. Depolanan bu verilerden faydalı bilgiler elde etmek ve bu bilgilerden en verimli şekilde yararlanabilmek ise başarı ve güç için en temel ihtiyaçlardır. Bu sebepten yaşadığımız çağ sıklıkla bilgi çağı olarak isimlendirilir.

İşletmeler için performans, ölçeklenebilirlik ve güvenilirlik her zamankinden daha fazla ön plana çıkmıştır. Bu önemli özelliklerin yanı sıra, son derece hızlı değişen pazar koşullarına ayak uydurabilecek rekabet avantajına erişmek için etkili ve çevik bir veri yönetim - analiz sisteminin işletmeler için önemi her geçen gün artmaktadır. Veri yönetimi ve analizi perspektifinden bakıldığında, birçok farklı noktadan gelen ham veriyi iş zekasına yönelik olarak biçimlendirebilmek ve Web ortamının sunduğu avantajlara erişebilmek de kritik öneme sahiptir.

Öneri sistemleri (ÖS) ise, işletmeler için kullanıcı veri tabanlarından değerli bilgiler çıkaran yeni bir teknolojidir. Kullanıcıların almayı düşündüğü ürünleri bulmasına yardım eden bu sistemler elektronik ticaret kavramına yeni bir boyut kazandırmıştır. Kullanıcıların satın almak istedikleri ürünlerin bulunması için öneri sistemleri veri analiz tekniklerini kullanırlar. Bunun için ya her bir ürüne bir tavsiye skoru üretirler yada en çok satılan n tane ürünün listesini kullanıcıya verirler. Örneğin ünlü bir kitap satış sitesi olan Amazon.com 'da kullanılan öneri sistemi, kullanıcının daha önceden hoşlandığını belirttiği kitaplara dayanarak yeni kitaplar tavsiye etmesidir.

Öneri sistemleri kullanıcıların demografiklerinden, en çok satılan ürünlerden kullanıcının geçmişteki alışveriş alışkanlıklarından veya kullanıcıların ürünler için yaptıkları değerlendirmelerden faydalanır. Temelde bu tekniklerin hepsi e-ticaret sitesini kullanıcı için cazip hale getirmeye böylece müşteri bağımlılığını temin etmeye çalışır. Müşteri bağımlılığını temin etmek e-ticaretin en temel stratejilerinden birisidir. Kısaca öneri sistemlerinden faydalanılarak siteye alışveriş için gelen

müşteriye özel dinamik sayfaların açılması, hangi ürünlerin kimler tarafında alındığı tespit edilerek özel promosyonların düzenlenmesi, müşterinin geçmiş değerlendirmelerinden yararlanarak önerilerin sunulması gibi yeni stratejiler belirlenerek hem kullanıcının istedikleri ürene daha kısa ve kolay şekilde ulaşması hem de elektronik ticaret yapan sitelerin satışlarının artırması sağlanabilir.

Bu çalışmada da yapay bağışıklık kullanılarak İşbirlikçi Filtreleme Tekniği tabanlı yeni bir öneri modeli sunulmuştur. Modelde kullanılmak üzere bir YBS algoritması olan aiNET seçilmiştir. Bunun sebebi bu algoritmanın yaygın olarak kullanılan bir öneri sistemi tekniği olan İşbirlikçi Filtreleme tekniğinin hem bilgi eksikliği hem de ölçeklenebilirlik problemlerini gidermede kullanılabileceğinin düşünülmesidir. Uygulamada veri kümesi olarak araştırmacılara açık olan ve yaygın olarak kullanılan MovieLens (ML) veri kümesi kullanılmıştır. Program kodları Delphi 7.0 kullanılarak yapılmıştır. Veri kümesine aiNET'den başka , k-ortalama algoritması ve standart işbirlikçi Filtreleme (SİF) yöntemi uygulanmıştır. Sunulan bu yeni modelin tanıtıldığı, parametre ve performans analizlerinin yapıldığı bu tez 6 ana bölümden oluşmaktadır.

Tezin ilk bölümünde yapılan çalışma tanıtılmış, amacı ve önemi anlatılmış ve geçmişte bu konuyla ilgili yapılmış çalışmalara ilişkin literatür özeti verilmiştir. İkinci bölümde öneri sistemleri ile ilgili bilgiler verilmiştir. Üçüncü bölümde ise insan bağışıklık sistemi ve bunu modelleyen yapay bağışıklık sisteminden bahsedilmiştir.

Tez uygulaması olarak sunulan yapay bağışıklık tabanlı İşbirlikçi Filtreleme Tekniğinin tanıtılması dördüncü bölümde yapılmıştır. Bu bölümde kullanılan veri tabanı, sunulan model, çalışma mantığı anlatılmış yapılan parametre ve performans analizlerinin sonuçları değerlendirilmiştir. Beşinci bölümde tezin genel bir değerlendirilmesinin yapıldığı sonuçlar kısmı bulunmaktadır. Son bölüm ise tez çalışması sırasında faydalanılan kaynakların bir listesini içermektedir.

1.1 Tezin Konusu, Amacı Ve Önemi

Günümüzde, veri yığınlarının çığ gibi büyüdüğü elektronik ortamlarda karar verirken kullanıcıya yardımcı olacak ve bu birikmiş verilerin içinde kaybolmasını engelleyecek bir sistemin ihtiyacı şiddetle hissedilmektedir. Bu amaçla özellikle alışveriş ve eğlence sitelerinde kullanıcının yolunu bulması için tasarlanmış İnternet tabanlı yazılım araçlarının yeni bir çeşidi olan “Öneri Sistemleri (*Recommender Systems*)” ön plana çıkmış araştırmacılar için cazip bir konu haline gelmiştir. “İşbirlikçi Filtreleme Tekniği (İFT -Collaborative Filtering Technique)” öneri sistemleri tarafından en yaygın olarak kullanılan tekniktir. İFT tabanlı öneri sistemleri ağırlıklı olarak benzer kullanıcıların komşuluklarını kullanarak işlem yaparlar. Yani müşterinin ilgilendiği ürünler hakkında diğer müşterilerin fikirlerine dayanarak bilgi elde etmeye çalışırlar. Kullanıcı benzerlikleri ise geçmişte farklı elemanlara verdikleri oylara bakılarak tespit edilir. Ortaya çıkışı 1990lara dayanan İşbirlikçi Filtreleme tekniğinin pek çok başarılı uygulaması olmasına rağmen halen giderilmesi gereken bazı problemler içermektedir. Bunlardan başlıcaları bilgi eksikliği (sparsity) ve ölçeklenebilirlik (scalability) problemleridir.

Bilgi eksikliği, sistemdeki ürünlerin sayısının arttıkça kullanıcının oyladığı ürünlerin oranının azalması ve buna bağlı olarak komşuluk hesabının zorlaşmasıdır. Ölçeklenebilirlik ise, genellikle büyük ölçekli veri tabanları için geçerli olan öneri sistemi algoritmalarının veri kümesinin büyüklüğünden dolayı çalışma zamanlarının uzaması performanslarının düşmesi olarak tanımlanabilir. Bu sorunların üstesinden gelmek ve öneri sistemlerinin verimliliği arttırmak için araştırmacıların yapmış oldukları çalışmalar bir sonraki bölümde özetlenmiştir. Bu tezde ise bu sorunları gidermeye ve böylelikle öneri sisteminin kalitesini ve performansını arttırmaya yönelik YBS tabanlı yeni bir İFT modeli sunulmuştur.

Yapılan çalışmada YBS kullanılmasının sebebi modellediği bağışıklık sisteminin az sayıda ve farklı gen kodlamalarına sahip bağışıklık elemanları ile çok sayıda ve farklı yapıdaki yabancı maddeleri tanıma özelliğine sahip olmasıdır. Ayrıca adaptif yapısı sayesinde bağışıklık sistemi vücuda daha önce hiç girmemiş ilk

defa karşılaştığı mikropları da tanıyabilir. Sunulan öneri modelinde veri tabanında bulunan kullanıcılar bağışıklık elemanlarını (antikorları) temsil eder ve oyladıkları filmler ile kodlanırlar. Sisteme yeni gelenler dışarıdaki yabancılar ise (antijenler) olarak kodlanır. Amaç mevcut antikorlardan, gelen antijenleri tanıyabilecek bir hafıza oluşturulmasıdır.

Bu tezin getirdiği yenilik YBS ile öneri sisteminin adaptif bir yapı kazanması, daha önce sisteme hiç gelmemiş kullanıcıları bile oluşturulan bu hafıza tabanlı öneri modeli ile sisteme en iyi şekilde adapte edilebilmesi ve bilgi eksikliğinden minimum derecede etkilenmesidir. Oluşturulan hafıza ayrıca tüm veri kümesinin saf bir örneğinin elde edilmesi anlamındadır. Böylece ölçeklenebilirlik probleminde de bir çözüm sunulmuş olunur.

Modelde kullanılmak üzere bir YBS algoritması olan aiNET seçilmiştir. Algoritma yaygın olarak kullanılan bir öneri sistemi tekniği olan İşbirlikçi Filtreleme algoritmasının hem bilgi eksikliği hem de ölçeklenebilirlik problemlerini gidermede kullanılabilir. Bilgi eksikliğini, içinde barındırdığı hiper mutasyon mekanizması ile daha önce hiç karşılaşılmamış ve komşuluk oluşturmak için yeterli bilgi olmayan bir antijen için en yakın komşuluklar oluşturulması sağlayarak giderir. Bunun için yeni bir mutasyon yöntemi kullanılması önerilmiştir. Bu yöntemde duyarlılığına göre seçilen bireyin rasgele bir bitine rasgele bir değer atamak yerine, saf bayes sınıflandırma tekniği kullanılarak olasılıksal bir değer atanmıştır. Böylelikle benzer kullanıcı olmasa bile oluşturulup hafızaya dahil edilmesi sağlanmıştır. Ayrıca aiNET'in içerdiği iki çeşit baskılama yöntemi vardır. Bu sayede de fazla veriler atılıp kümenin bir örneğinin elde edilmesi verimli bir şekilde sağlanır. Bu da ölçeklenebilirliği sağlamak için hedeflenen esas noktadır. Geliştirilen uygulamanın sonucuna bakıldığında ise uygun parametre değerleri seçildiğinde modelin diğer yöntemler ile benzer ve beklenen hata seviyesinde tahmin ürettiği gözlemlenmiştir. Buna karşın yapılan tavsiyenin doğruluğunda ve hızında sağladığı artış sunulan modelin kullanımının daha avantajlı olduğunu göstermiştir.

1.2 Literatür Özeti

Günümüzde sayısız sayıda firma interneti pazarlama ve ticaret için kullanmaktadır. E-ticarete olan bu yönelim, firmaların müşteriye daha çok ürün seçeneği sunmasına izin verir. Ancak seçeneklerin bu kadar artması kullanıcılarında tercih yapmasını güçleştirmiştir. Bu sorunun üstesinden gelmek için pek çok öneri tekniği geliştirilmiştir. Bunlardan en popülerleri film, makale, ürün web sayfası tavsiye etmek gibi pek çok uygulamada kullanılan işbirlikçi filtreleme tekniğidir (İFT). Bu tekniğin temelleri ilk olarak 1994'de Resnick ve ark. tarafından atılmıştır. Bu makalede haber gruplarında kullanıcıya haber önermek İFT tabanlı bir öneri sistemi sunulmuştur. İFT' de iki kullanıcı arasındaki benzerliğin hesaplanması için sıklıkla kullanılan Pearson Bağıntısı ilk defa literatürde bu makalede geçmiştir.

1998 yılında Breese ve ark. tarafından İFT için kullanılan bağıntı katsayısı, vektör benzerliği ve istatistiksel Bayesian tabanlı metotlar birbirleri ile kıyaslanmış, bu metotların doğru tahmin üretme yetenekleri aynı problemler üzerinde ölçülmüştür. Deneysel üç veri kümesi üzerinde, dört çeşit deneyde 2 değerlendirme ölçütü kullanılarak gerçekleştirilmiştir. Sonuçlara göre Bağıntı ve Bayesian tabanlı metotlar vektör benzerliği metodundan daha iyi sonuç vermiştir. Bu iki metot arasında tercih yaparken de veri tabanının ve uygulamanın yapısının göz önünde bulundurulması gerektiği belirtilmiştir.

2000 yılında Sarwar ve arkadaşlarının yaptığı çalışmada İFT tabanlı öneri sistemlerinin üç aşamada müşteriye tavsiye verdiği belirtilmiştir. Bunlardan ilki kullanıcının oyladığı ürünlere bakılarak kullanıcı profilinin oluşturulmasıdır. İkinci aşama sistemin makine öğrenmesi veya istatistiksel teknikler kullanarak benzer davranışlara sahip komşuluklar olarak adlandırılan kullanıcı kümelerini inşa etmesidir. Son aşama ise tahmin ve tavsiye hesaplama sürecidir.

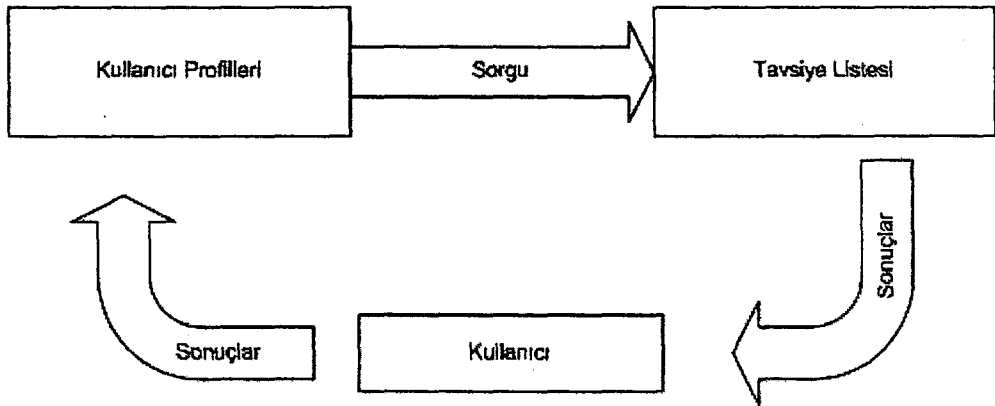
Ancak e-ticaret sitelerindeki müşteri ve ürün sayısının hızla artması bilgi eksikliği ve ölçeklenebilirlik problemlerinin ortaya çıkmasına sebep olmuştur. Bu

problemlerin çözümü için 1998'den beri pek çok çalışma yapılmıştır. Bu tezde ise bu sorunlara evrimsel bir algoritma olan YBS kullanılarak bir çözüm önerilmiştir.

YBS 1990'dan itibaren gelişmesi hız kazanmış, insan bağışıklık sistemini modelleyen evrimsel bir algoritmadır. Öneri sistemlerinde kullanılmaya başlanması ise 2002 yılında Cazyer ve Aickelin yaptığı çalışmaya dayanır. Bu çalışmada bağışıklık sisteminin dağıtık adaptif ve kendi kendini organize edebilen karmaşık biyolojik yapısını modelleyen yapay bağışıklık sistemi ile bir film tavsiye işlemi gerçekleştirmeyi amaçlamaktadır. Hipotezlerine göre YBS bu hedeflerini başarmaya aday olabilecek iki esas görüş üzerine kurulmuştur. Bunlar eşleme mantığı için antijen-antikor kesişmesi ve çeşitlilik için antikor-antikor kesişmesidir. Çalışmanın sonucunda deneysel sonuçlar verilmiş ve bunlar diğer tavsiye yöntemleri ile kıyaslanmıştır. Chen ve Aickelin'in 2004'de yaptığı çalışmada YBS kullanılarak bir film tavsiye etmek için İFT tabanlı bir öneri sistemi geliştirilmiştir. Bu çalışmada duyarlılık olgunlaşması sırasında kullanılabilinecek kendall tau ve ağırlıklı kappa benzerlik ölçütleri kıyaslanmış ve ağırlıklı kappa'nın daha verimli sonuçlar verdiği kanısına varılmıştır.

2. ÖNERİ SİSTEMLERİ

Günlük hayatta, sıklıkla fazla bilgi ve deneyime sahip olunmayan konular hakkında karar verilmesi gerekir. Bu durumlarda genellikle daha deneyimli kişilere danışılarak fikirleri alınır ve elde edilen ön verilere dayanarak daha sağlıklı kararlar alınabilir. Veri yığınlarının çığ gibi büyüdüğü elektronik ortamlarda ise karar verirken kullanıcıya yardımcı olacak ve bu birikmiş verilerin içinde kaybolmasını engelleyecek bir sistemin ihtiyacı şiddetle hissedilir. Bu amaçla özellikle alış verişi ve eğlence sitelerinde kullanıcının yolunu bulması için tasarlanmış Internet tabanlı yazılım araçlarının yeni bir çeşidi olan “Öneri Sistemleri - ÖS (*Recommender Systems*)” ön plana çıkar. Öneri sistemlerinin, ürünler hakkında müşterilere ilgileri doğrultusunda farklı taleplerine cevap verebilecek nitelikte olan tutarlı tavsiyeleri bilgi eksikliğini en iyi şekilde gidererek kısa sürede üretmesi beklenir. Şekil 2.1 ‘de öneri verme süreci gösterilmiştir. Şekilde de gösterildiği gibi öncelikle kullanıcıdan tercihler alınarak, kullanıcı profili oluşturulur. Diğer kullanıcılar ile olan benzerliğin bulunması için bir sorgu sürecinden geçilir ve sorgunun sonucuna dayalı olarak tavsiye sonuçları kullanıcıya döndürülür. Kullanıcıdan gelen geri besleme ile de sistemin yapısı sağlamlaştırılır. Sağlanan döngü ile tavsiyelerin kalitesi ve doğruluğu artırılır.



Şekil 2.1: Öneri Verme Süreci

Bu bölümün devamında sırasıyla öneri sistemlerinin yapısından, işbirlikçi filtreleme tekniğinden ve öneri sistemlerinin eksik yönlerinden bahsedilecektir.

2.1 Öneri Sistemlerinin Yapısı

İçerik Tabanlı Filtreleme (Content based filtering) ve İşbirlikçi Filtreleme (Collaborative Filtering) olmak üzere öneri sistemlerinin kullandığı iki temel filtreleme tekniği vardır. İçerik tabanlı filtreleme teknikleri kullanıcı tercihlerine ait bir profil oluşturmak için oylanmış olan bilgi kaynağının içeriğini analiz eder. Oluşturulan bu profil daha önceden görülmemiş diğer bilgi kaynaklarını oylamak için veya arama motoru için bir sorgu oluşturmakta kullanılabilir (Pazzini M.J.1999). İşbirlikçi filtreleme teknikleri ise içerik hakkında herhangi bir bilgi gerektirmez. Kullanıcıların geçmişte verdikleri oyların benzerliklerine bakarak komşulukları tespit eder ve bunlara dayanarak tavsiye üretir.

2.1.1 İçerik Tabanlı Filtreleme Teknikleri

Otuz yılı aşkın süredir bilgisayar bilimcileri yazılım teknolojilerini kullanarak hızla biriken verileri tanıma ve sınıflandırma problemini çözmeye çalışmaktadırlar. Bu amaç için geliştirilen yazılımlar sayesinde otomatik olarak her bir elemanın içeriğinin tanımı üretilir daha sonra bu tanımlar ile kullanıcının ihtiyacı olan elemanın tanımı kıyaslanır. Kullanıcının ihtiyaç duyduğu tanım ise ya bir sorgu ile kullanıcıdan alınır ya da kullanıcının daha önce ilgilendiği elemanların gözlemlenmesi ile öğrenilir. Bahsedilen bu teknikler içerik tabanlı filtreleme olarak adlandırılır çünkü filtreleme işlemi elemanların içeriği analiz edilerek gerçekleştirilir (Herlocker J. 2000).

Bu filtreleme teknikleri genellikle kullanıcıya tercihlerine göre doküman tavsiyesi yapılırken kullanılır. Kullanıcının oyladığı elemanın (makaleler, haber gruplarındaki mesajlar vb.) içeriği ile tavsiye edilmesi düşünülen elemanın içeriğini analiz ederek tavsiye üretir. Tavsiye verilirken temel alınacak olan içeriğinin düzenini bulması için dokümanın içeriğini analiz eden pek çok algoritma geliştirilmiştir. Bu algoritmaların çoğu dokümanın hangi sınıfa (hoşlanılan veya hoşlanılmayan) ait olduğunu öğrenmeye çalışan sınıflandırma algoritmalarının

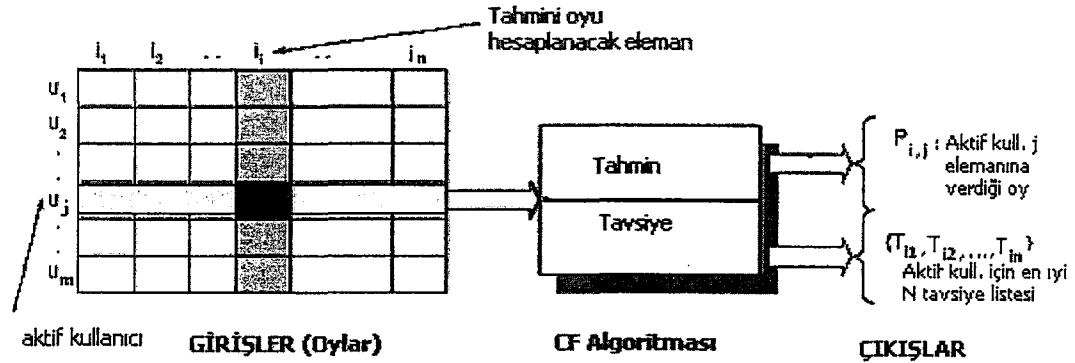
özelleştirilmiş versiyonlarıdır. Geriye kalan diğer algoritmalar ise dokümana verilecek (oy gibi) sayısal bir değeri tahmin etmeye çalışırlar (Pazzini M.J. 1999).

İçerik tabanlı filtreleme sistemi tasarlanırken iki tane temel problemin çözülmesi gerekir. Bunlardan ilki bir dokümanın nasıl temsil edileceğini belirlemektir. Hemen hemen tüm içerik tabanlı filtreleme sistemini kullanan yaklaşımlar dokümanı göstermek için dokümanın içinde geçen “önemli” kelimeleri kullanırlar. Hangi kelimelerin önemli olduğuna karar vermek içinse literatürdeki çeşitli yöntemlerden biri kullanılarak kelimelerin ağırlıkları hesaplanır ve kullanılan yöntemle göre en yüksek veya en düşük ağırlık değerine sahip ilk n kelime dokümanı göstermek için seçilir. Bu yöntemlerden birisi terim frekans indekslemesidir. İkinci sorun ise henüz görülmemiş dokümanların tavsiye edilmesini mümkün kılan bir modelin oluşturulmasıdır (Pazzini M.J. 1999). Dokümanın hangi kelimeler ile temsil edileceği belirlendikten sonra “karar ağaçları”, “bayes sınıflandırıcı” veya “winnow” gibi bir sınıflandırma algoritması kullanılarak dokümanlar gruplanabilir (Blum A. 1997).

2.1.2 İşbirlikçi Filtreleme Teknikleri

İşbirlikçi filtreleme teknikleri (İFT) öneri sistemleri tarafından en yaygın olarak kullanılan tekniklerdendir. İFT hiç tanımadığımız kişilere bilgisayar yardımıyla tavsiye vermemize yardım eder. Başka bir deyişle İFT sistemleri bilgisayarın ve insanın iyi yapabildiği şeyi yapmasına izin vererek ortak çalışma imkanı sunar. Şöyle ki; kullanıcı elemanları okumada ve ya değerlendirmede iyidir bilgisayar ise bu değerlendirmeler arasındaki benzerliklerin bulunması için gerekli olan hesaplamaları yapmakta. Kullanıcıların bir eleman için yaptıkları değerlendirmeye “oylama” denir. Oylama genellikle 1-5 veya 1-7 gibi belli bir aralıkta yapılır ve kullanıcının o eleman hakkındaki düşüncesinin iyimi kötümü olduğunu gösterir. Bilgisayarın rolü ise bir kullanıcının henüz görmediği bir elemana vereceği oyu önceden tahmin etmektir. Tahmin hesaplamasındaki ilk adım elemanları benzer şekilde oylayan kullanıcı gruplarının tespit edilmesidir. Bu kullanıcı grubuna “komşuluk” adı verilir ve bir kullanıcının bir eleman için yaptığı

tahmin, o eleman için komşularının yaptığı oylamalara bakılarak hesaplanır. Buradaki esas düşünce eğer bir kullanıcı geçmişte komşuları ile aynı fikirde ise gelecekte de aynı fikirde olma olasılığının yüksek olmasıdır (Miller B. N. 2003).



Şekil 2.2: İşbirlikçi filtreleme süreci (Sarwar B.M. ve ark. 2001)

Şekil 2.2 İFT sürecinin şematik yapısını gösterir. Görüldüğü üzere $m \times n$ boyutlu oy bilgisini içeren bir kullanıcı-ürün matrisi giriş olarak alınmaktadır. Matristeki her bir $a_{i,j}$ elemanı i . kullanıcının j . ürüne verdiği oyu göstermektedir. Bu matrise İFT algoritması uygulanır ve çıkış olarak da tahmin ve tavsiyeler elde edilir.

İşbirlikçi filtreleme tekniğinin aktif kullanıcı için veri tabanında bulunan diğer benzer kullanıcıların oylarına dayanarak faydalı tahminler yapma işlemi olduğu yukarıda belirtilmişti. Bu filtreleme tekniği kendi içinde iki sınıfta incelemek mümkündür: Hafıza tabanlı İFT ve Model tabanlı İFT. (Breese J.S. ve ark. 1998)

Hafıza tabanlı İFT tahmin yapmak için veritabanı üzerindeki tüm kullanıcı oyları üzerinde ayrı ayrı işlem yaparken Model tabanlı İFT bunun aksine kullanıcı veritabanını bir model oluşturmak veya öğrenmek için kullanır ve daha sonra tahmin yapmak için bu modelden faydalanır. (Breese J.S. ve ark. 1998)

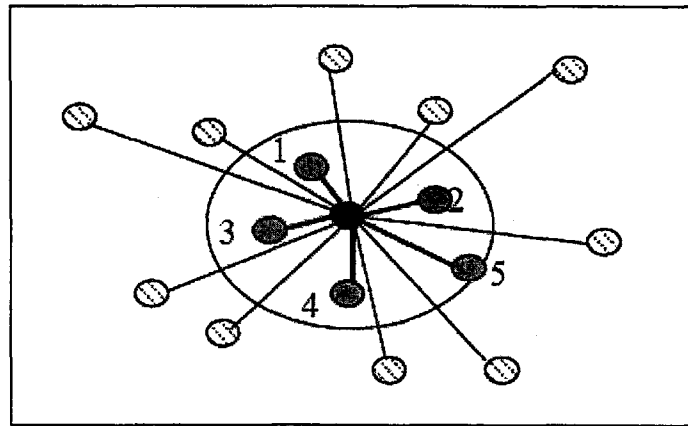
İFT' yi sınıflandırmak için kullanabileceğimiz bir diğer yaklaşımda hesaplamada kullanılan oyların çeşididir. Gerçek olmayan oylar (implicit voting) ve gerçek oylar (explicit voting) olmak üzere iki tür oydan söz edilebilir. Açık oylama bir kullanıcının bir eleman üzerinde bilinçli şekilde yaptığı veri tabanına kaydı

bulunan oylamadadır. Gerçek olmayan oylar ise veritabanındaki bilgi eksikliğini tamamlamak için kullanıcının geçmişteki seçimlerine, davranışlarına, verdiği oylara bakılarak yapay olarak üretilen, veritabanına sonradan kaydedilen oylar kastedilir. Gerçek olmayan oylar web uygulamalarında tarayıcı verilerine, satış geçmişine veya diğer veri erişim örüntülerine dayandırılabilir (Breese J.S. ve ark. 1998). Yapay oyların elde edilmesinde veri madenciliği teknikleri veya yapay zeka algoritmaları kullanılabilir.

Bu bölümün geri kalan kısmında sırasıyla komşulukların oluşturulmasından, bunlara dayalı tahmin ve tavsiyelerin üretilmesinden ve üretilen tahmin ve tavsiyelerin doğrularının değerlendirilmesi için kullanılan değerlendirme tekniklerinden bahsedilecektir.

2.1.2.1 Komşulukların Oluşturulması ve Seçimi

Kullanıcıya belli bir ürün için tavsiye üretirken yapılması gereken ilk işlem kullanıcı komşuluklarının oluşturulmasıdır. Komşuluklar kullanıcıların geçmişte elemanlara verdikleri oyların benzerliklerine bakılarak tespit edilir. Benzerliklerin tespitinde istatistiksel yöntemlerden veya makine öğrenmesi tekniklerinden faydalanılabilir. Buradaki esas düşünce eğer bir kullanıcı geçmişte komşuları ile aynı fikirde ise gelecekte de aynı fikirde olma olasılığının yüksek olmasıdır.



Şekil 2.3 : Komşuluk Gösterimi. Aktif kullanıcı ile diğer kullanıcılar arasındaki uzaklık gösterilmiş ve en yakın k (örnek için 5) tane kullanıcı komşu olarak belirlenmiştir (Sarwar B.M. ve ark. 2000).

Şekil 2.3' de iki boyutlu bir uzayda basitçe en yakın komşuların bulunması gösterilmiştir. Çemberin içinde kalan 1,2,3,4 ve 5 numaralı kullanıcılar komşu olarak seçilip tahmin ve tavsiye üretirken kullanılacaktır.

Kullanıcılar ile aktif kullanıcı arasındaki benzerliğin hesaplanmasında sıklıkla kullanılan yöntemlerin başında "Pearson Bağıntısı (Pearson Correlation)" gelir. Pearson bağıntısı ilk olarak 1994 yılında GroupLens projesinde Resnick ve ark. tarafından kullanılmıştır (Breese J.S. ve ark. 1998). Pearson bağıntısının hesaplanması sonucunda elde edilen değer ne kadar yüksek ise kullanıcıların benzerliği o kadar fazladır. Matematiksel olarak Pearson bağıntısı Formül 2.1'de ifade edildiği gibidir.

$$\text{Benzerlik}(a,k) = \frac{\sum_{i=1}^n (a_i - \bar{a})(k_i - \bar{k})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2 \sum_{i=1}^n (k_i - \bar{k})^2}} \dots\dots\dots(2.1)$$

a_i = aktif kullanıcının i filmine verdiği oy

k_i = k kullanıcısının i filmine verdiği oy

n = a ve k kullanıcısının ortak oyladıkları film sayısı

\bar{a}, \bar{k} = sırasıyla a ve k kullanıcısının tüm filmlere verdikleri oyların ortalaması

Bir kullanıcının tüm filmlere verdiği oyların ortalamasının hesaplamak için eşitlik 2.2 'deki formül kullanılır. Buradaki I_a terimi a kullanıcısının oyladığı filmlerin kümesidir. (Breese J.S. ve ark. 1998)

$$\bar{a} = \frac{1}{|I_a|} \sum_{j \in I_a} a_j \dots\dots\dots(2.2)$$

Pearson bağıntı hesabı için bazı istisna durumlar söz konusudur. Bunlardan ilki iki kullanıcısının oyladıkları hiç ortak film bulunmamasıdır ($n=0$). Bu durumda Benzerlik(a,k) sıfır olarak kabul edilir. İkinci durum ise paydanın değerinin sıfır çıkmasıdır. Bu sıfıra bölüm hatasına sebep olacağından Benzerlik(a,k)'nin değeri sıfır olarak atanır. Diğer bir durum ise ortak oylanan film sayısının istenen sayıdan (p) düşük olmasıdır. Yani ($n<p$) gibi bir durumda Benzerlik(a,k)'nin değerini direkt sıfıra eşitlemek hatalı olur. Sayısı az da olsa ortak oylanan film vardır. Bu durumda Benzerlik(a,k)= (n/p) *Benzerlik(a,k) olarak hesaplanabilir. (Cazyer S., Aickelin U. 2002)

Kullanıcılar arasındaki benzerliğin bulunması amacı ile kullanılan diğer bir yöntem ise “Vektör Benzerliği (vector similarity)”dir. Bu yöntem sıklıkla bilgi erişim (information retrieval) alanında kullanılır. Her bir doküman için kelime frekanslarının tutulduğu vektörler tanımlanır ve bu vektörler arasındaki benzerlik aralarındaki kosinüs açısı hesaplanarak ölçülür. Bu yaklaşımı işbirlikçi filtreleme tekniğine uygulanması mümkündür. Kullanıcılar başlıkta ve dokümanda geçen kelimeler olarak düşünülürken oylar ise bu kelimelerin frekansının yerine geçer. Bu algorithmada gözlemlenen tüm oyların değerinin pozitif olduğu düşünülür negatif oylama söz konusu değildir. Eğer bir eleman için herhangi bir oy gözlemlenmemiş ise onun değeri de sıfır olarak kabul edilir. Matematiksel olarak ifadesi (2.3) gösterildiği şekildedir.

$$\text{benzerlik}(a,k) = \sum_i \frac{a_i}{\sqrt{\sum_{j \in I_a} (a_j)^2}} \frac{k_i}{\sqrt{\sum_{j \in I_k} (k_j)^2}} \dots\dots\dots (2.3)$$

2.3'deki formülün paydasında bulunan kareli terimler, oyların normalizasyonunu sağlamak içindir. Böylelikle ilk etapta pek çok oylanmamış filmi bulunan kullanıcıların birbirleri ile yüksek benzerlik göstermesine mani olunur. Başka normalizasyon teknikleri de kullanılması mümkündür.

Breese ve ark. (1998) kullanıcıların benzerlikleri hesaplanırken kullanılan Pearson bağıntısının performans açısından vektör benzerlik yönteminden daha üstün olduğunu göstermişlerdir.

Yukarıdaki yöntemler kullanılarak benzerlikler hesaplandıktan sonra sıra hangi kullanıcıların komşu olarak seçileceğine gelir. Bunun için kullanılan başlıca yöntemler “Bağıntı Eşiği (Correlation Thresholding)” ve “n en iyi komşu tekniği (best n neighbors)” dir. Bağıntı eşiği yönteminde, verilen bir eşik değerinden yüksek benzerliğe sahip kullanıcılar komşu olarak seçilir. n en iyi komşu tekniğinde ise en iyi benzerlik değerine sahip n tane kullanıcı komşu olarak seçilir. İkinci yöntemin performans açısından daha üstün olduğu gösterilmiştir (Herlocker ve ark. 1999). Kullanıcıya en yakın n tane komşunun seçilmesi yerine önerilen diğer bir yaklaşım ise komşuluk kümesinin merkezine en yakın n tane komşunun seçilmesidir. Şekil 2-4’de şematik olarak bu iki yöntem gösterilmiştir (Sarwar B.M. 2001).



Şekil 2.4 : (a): kullanıcıya en yakın n tane kullanıcının komşu olarak seçilmesi.

(b): küme merkezine en yakın n tane kullanıcının komşu olarak seçilmesi.

Komşuluk seçimi için önerilen bir diğer yöntemde ise yapay bağımsızlık sisteminden esinlenilmiştir. Kim için tahmin veya tavsiye üretilecekse o kullanıcı antijen olarak düşünülür. Diğer kullanıcılar ise antijeni tanıması gereken antikordlardır. Antikorum yoğunluğu, antijen ile yani aktif kullanıcı ile eşleşmesine bağlı olarak artarken diğer antikorlarla eşleşmesine bağlı olarak azalır. Yoğunluğu belli bir değer altına düşen antikorlar elenirken, yüksek yoğunluğa sahip olanların sayısı artırılır. Antikorların yoğunluğunda 10 iterasyon boyunca hiçbir değişiklik olmazsa veya antikor yoğunluğu maksimum. değerine ulaşırsa algoritma sonlandırılır ve havuzda bulunan antikorlar komşu olarak seçilir. Bu yaklaşım YBS’ de jerne’nin Idiotipic ağ yapısından uyarlanmıştır ve adaptif komşu seçimine olanak tanır. Her bir

adımdaki antijen ve antikorların yoğunluğu eşitlik 2.4 tarafından kontrol edilir (Cazyer S., Aickelin U. 2002).

$$\begin{aligned} \frac{dx_a}{dt} &= (\text{AntijenTeş viği}) - (\text{AntikorBas kılanması}) - (\text{ÖlümOranı}) \\ &= k_1 m_a x_a y - \frac{k_2}{N} \sum_{k=1}^N m_{ak} x_a x_k - k_3 x_a \dots \dots \dots (2.4) \end{aligned}$$

$m_{a,k}$ = | benzerlik(a,k) |

k_1 = Teşvik Sabiti

k_2 = Baskılama Sabiti

k_3 = Ölüm oranı Sabiti (0.1 olarak kabul edilmiş)

N = Maksimum antikor Sayısı (100 olarak kabul edilmiş)

x_a = Havuzda o anda bulunan antikor yoğunluğu (Başlangıç değeri olarak 10 kabul edilmiş. Yoğunluk 0-100 aralığında değişebilir.)

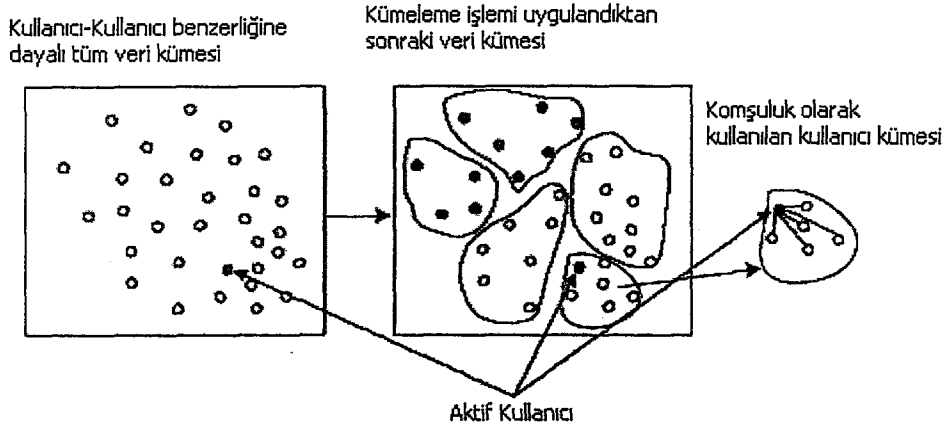
y = Antijen yoğunluğu (Her zaman 1'e eşit)

Çizelge 2.1 : Kümeleme ile komşu seçimini kullanan İFT algoritması

Kümeleme yardımı ile komşu seçiminin algoritması aşağıdaki gibidir :
1. Kümeleme algoritması kullanılarak eğitim kümesi p parçaya ayrılması: A eğitim kümesi $A_i \cap A_j = \emptyset$, $1 \leq i, j \leq p$ ve $A_1 \cup A_2 \cup \dots \cup A_p = A$ olacak şekilde A_1, A_2, \dots, A_p gibi p parçaya bölünür.
2. Verilen u kullanıcısı için komşuluklar belirlenmesi: Eğer $u \in A_i$ ise, A_i kümesi u kullanıcısı için komşuluk olarak atanır.
3. Komşuluklar elde edildikten sonra, klasik İFT kullanılarak tahmin ve tavsiyelerin üretilmesi.

Çok büyük veri tabanları söz konusu olduğunda , kullanıcının komşuluklarını bulmak için veri tabanındaki tüm elemanlar ile benzerliğinin hesaplanması oldukça pahalı bir işlemdir. Bu sorunu aşmak için sunulan bir yaklaşım veri tabanına kümeleme tekniklerinden birisi uygulamak ve benzer kullanıcı grupları oluşturmaktır. Böylece kullanıcının komşulukları hesaplanırken tüm veri tabanındaki kullanıcıların kullanılması yerine sadece kendi kümesi içinde kalan kullanıcıların

kullanılması sağlanmış olur. Şekil 2.5’de bu fikrin yapısı, çizelge 2.1’de ise algoritması verilmiştir (Sarwar B.M ve ark. 2002).



Şekil 2.5 : Kümeleme yardımı ile komşu seçimi

Komşu seçimin bu şekilde yapılmasının iki önemli avantajı vardır. Bunlardan ilki büyük veritabanının ölçeklenebilirliğini artırır. Buna bağlı olarak elde edilen ikinci avantaj ise komşular bulunurken ve tahmin hesaplanırken geçen sürenin oldukça azalmasıdır.

2.1.2.2 Tahmin ve Tavsiye Üretme

Bir önceki bölümde belirtilen bir yöntem kullanılarak komşuluklar oluşturulup komşular seçildikten sonra sıra asıl amaca yani kullanıcının isteklerine cevap verebilecek ve onun doğru yönlendirilmesini temin edebilecek nitelikte tahmin ve tavsiye üretmeye gelmiştir.

Tahmin ve tavsiye temelde birbirlerine çok yakın ve birbirlerini tamamlayan iki kavram olmasına rağmen özünde taşıdıkları anlamlar farklıdır. Tahmin üretmek, adı üstünde kullanıcının eleman için yapacağı değerlendirmeyi önceden tahmin etme işlemidir. Tavsiye ise kullanıcının yüksek oy vereceği elemanın önceden bulunup kullanıcıya önerilmesidir. Tavsiye üretmek öneri sistemlerinin doğasına daha uygun olmasına rağmen yapılan tahminin doğruluğunun ölçülmesi nicel olarak tavsiyenin

doğruluğunun ölçülmesinden daha kolaydır. Tahmin ve tavsiyelerinin doğruluğunun değerlendirilmesi bir öneri sisteminin ne kadar sağlıklı çalıştığını bulma hususunda önem kazanmaktadır. Bir sonraki bölümde değerlendirme ölçütleri daha ayrıntılı şekilde incelenecektir. Bu bölümün geri kalan kısmında ise tahmin ve tavsiye üretmede kullanılan yöntemler açıklanacaktır.

a gibi bir kullanıcının j elemanı için yapacağı tahmini hesaplamak için kullanılan en genel yöntem, a kullanıcısının komşularının j elemanına verdiği oyların kullanılmasıdır. Bu yöntemin formülü eşitlik 2.5'de gösterildiği gibidir. (Resnick P. 1994)

$$P_{a,j} = \bar{a} + \frac{\sum_{(k \in N) \wedge (k_oylar_j)} (k_j - \bar{k}) * benzerlik(a,k)}{\sum_{k \in N \wedge (k_oylar_j)} |benzerlik(a,k)|} \dots\dots\dots(2.5)$$

\bar{a} = a kullanıcısının verdiği oyların ortalaması

k_j = k kullanıcısının j elemanına verdiği oy

\bar{k} = k kullanıcısının verdiği oyların ortalaması

$benzerlik(a,k)$ = a ile k kullanıcısı arasındaki benzerlik

Bu formül ile tahmin hesaplanırken dikkat edilmesi gereken önemli bir nokta, hakkında tahmin yapılacak elemanı oylamamış yüksek benzerlik değerine sahip komşuların olabilmesidir. Onlar hesaplamalara komşu olarak dâhil edilmemelidir.

İFT tabanlı tavsiye sistemlerinin son adımı ve esas amacı kullanıcının komşularını kullanarak en iyi N tane tavsiye üretmektir. (Kullanıcının alması en muhtemel ilk 10 ürünün listelenmesi gibi.) Bunun için öncelikle K komşuluk kümesi araştırılır ve komşuların değerlendirdiği elemanlara bir oylama sayacı atanır. Tüm elemanlar için oylama sayacı hesaplandıktan sonra, elemanlar önce aldıkları oylara büyükten küçüğe göre sıralanırlar. Daha sonrada sıralan bu elemanlar oylama sayacına göre büyükten küçüğe doğru kendi içinde tekrar sıralanırlar. İlk N sıradaki eleman kullanıcıya tavsiye olarak verilir (Sarwar B.M. 2001).

2.1.2.3 Değerlendirme Ölçütleri

Öneri sistemlerinde tahmin veya tavsiyenin doğruluğunun araştırılması için kullanılan pek çok değerlendirme ölçütü mevcuttur. Bir önceki bölümde açıklandığı gibi tahmin ve tavsiyenin elde ediliş biçimleri birbirlerinden farklıdır bu da doğruluklarını araştırırken farklı değerlendirme ölçütleri kullanılmayı gerektirir. Tahmin ve tavsiyelerin doğruluğundan başka sistem değerlendirilirken kullanılan diğer bir ölçüt ise sistemin performansıdır. Bunun için sistemin ne kadar sürede kaç tane tavsiye ürettiği gibi soruların cevapları aranır. Bir sonraki bölümde öncelikle tahmin ve tavsiye değerlendirme ölçütleri ele alınacaktır. Daha sonra ise performans değerlendirme kriterleri incelenecektir.

2.1.2.3.1 Tahmin Değerlendirme Ölçütleri

Öneri sistemlerinin temel amacı, kullanıcı istediklerinin en etkili şekilde belirlenmesi olduğundan yapılan tahmin doğruluk ve kapsam olmak üzere iki farklı açıdan değerlendirilir.

a) Tahminin Doğruluğu :

Literatürde tahmin doğruluğunun ölçülmesi için pek çok yöntem önerilmiştir. Bunları “istatistiksel doğruluk ölçütleri” ve “karar destek doğruluk ölçütleri” olmak üzere iki başlık altında toplayarak inceleyebiliriz. (Sarwar B.M. 2001)

İstatistiksel Doğruluk Ölçütleri :

Sistemin ürettiği sayısal tahminler ile kullanıcı tarafından verilmiş sayısal değerlendirmeler arasındaki yakınlığı ölçer. Kullanılan genel ölçütler şunlardır:

a- Mutlak Hatanın Ortalaması –MHO (Mean Absolute Error - MAE) : Bir ürün için kullanıcının verdiği oyu ne kadar doğru tahmin edebildiğimizi ölçer. P vektörü, bir kullanıcının N tane ürün için vereceği oyların hesaplanan tahmini değerini tutsun. R vektörü ise kullanıcının N tane ürüne verdiği gerçek oyları içersin.

$|p_i - r_i|$ ifadesi bir eleman için gerçek oy ile tahmin edilen oy arasındaki mutlak hatayı verir. MHO öncelikle N tane ürün için mutlak hataların toplamını bulur ve ortalamasını hesaplar. Hesaplanışı eşitlik 2.6' de gösterildiği gibidir.

$$MHO = \frac{\sum_{i=1}^N |p_i - r_i|}{N} \dots\dots\dots(2.6)$$

b- Hata Karelerinin Ortalamasının Karekökü –HKOK (Root Mean Squared Error-RMSE): Burada da P vektörü , bir kullanıcının N tane ürün için vereceği oyların hesaplanan tahmini değerini tutarken R vektörü ise kullanıcının N tane ürüne verdiği gerçek oyları içerir. HKOK büyük hatalara küçük hatalara göre daha fazla ağırlık verme eğilimdedir. Hesaplanışı eşitlik 2.7' de gösterildiği gibidir.

$$HKOK = \sqrt{\frac{\sum_{i=1}^N (p_i - r_i)^2}{N}} \dots\dots\dots(2.7)$$

c- Yüzdelerik Mutlak Hatanın Ortalaması – YMHO (Mean Absolute Percentage Error-MAPE): Amacı yukarıda açıklanan iki yöntem ile aynıdır. Sadece formülasyonu değişiktir. Mutlak hatanın ortalaması yerine yüzdelerik değerlerinin ortalaması kullanılır. Hesaplanışı eşitlik 2.8' de gösterildiği gibidir.

$$YMHO = \frac{\sum_{i=1}^N \frac{|p_i - r_i|}{p_i} * 100}{N} \dots\dots\dots(2.8)$$

d- Bağıntı (Correlation) : Bağntı iki veri vektörü arasındaki uyuşmanın istatistiksel olarak ölçümüdür. İki vektör arasındaki doğrusal uyuşmanın ölçülmesi için genellikle standart pearson bağıntısı kullanılır. Pearson bağıntısının hesaplanışı 2.9'da gösterildiği gibidir.

$$Bağağın_{PR} = \frac{\sum_{i=1}^N (p_i - \bar{p})(r_i - \bar{r})}{N * \sigma_p * \sigma_R} \dots\dots\dots(2.9)$$

Sırasıyla p_i ve r_i hesaplanan tahmini oy ve verilen gerçek oyu gösterir. \bar{p} ve \bar{r} P ve R vektörlerinde yer alan elamanların ortalamasıdır. σ_p ve σ_R ise kümelerin standart sapmasını gösterir. σ_p standart sapma değerinin hesaplanması eşitlik 2.10' de verilmiştir.

$$\sigma_p = \sqrt{\sum_{i=1}^N (p_i - \bar{p})^2} \dots\dots\dots(2.10)$$

Yukarıda bahsedilen ilk üç yöntemde hesaplanan değer ne kadar düşük ise sistemin ürettiği tahminlerin doğruluğun o kadar fazla olduğu söylenir. Son yöntemde ise bunun tersine bağıntı değeri ne kadar büyük olursa tahminin o kadar doğru olduğu söylenir.

Karar Destek Doğruluk Ölçütleri:

Karar destek doğruluk ölçütleri, üretilen tahminlerin kullanıcıya eleman seçiminde ne kadar faydalı olduğunu ölçer. Bu değerlendirme ölçütleri gözlemlere dayanır, şöyle ki pek çok kullanıcı için tahmin üretmek için kullanılan filtreleme işlemi ikili bir işlemden ibarettir. Mesela bir kullanıcı bir filmi ya izleyecektir yada izlemeyecektir. Eğer kullanıcı sadece 4'un üzerinde oy alan filmleri izliyor ise bu durumda bir filmin 1 ile 3 arasında oy almış olması kullanıcıyı ilgilendirmiyor demektir (Miller B. N. 2003). En yaygın kullanılan karar destek ölçütleri ROC ve PRC duyarlılıklarıdır. Bunları sırası ile açıklarsak;

a- ROC Duyarlılık Ölçüsü : ROC duyarlılığı filtreleme sisteminin doğru tanı koyabilmesinin ölçüsüdür. Yani, sistemin kullanıcıyı düşük oy alan elemanlardan uzak tutmada ve yüksek oy alanlara yönlendirmede ne kadar başarılı olduğunun bir göstergesidir. İşlemsel olarak ise ROC eğrisinin altında kalan alandır. ROC eğrisi alanı X ayırıcı değişkenine göre *duyarlılık (sensitivity)* ve *belirlilik (specificity)*

olmak üzere iki parçaya böler. Duyarlılık, rasgele seçilen iyi bir elemanın filtre tarafında kabul edilme olasılığıdır. Belirlilik ise rasgele seçilen kötü bir elemanın reddedilme olasılığıdır. Eğri 0-1 aralığında çizilir. Eğri altındaki alan filtre ne kadar çok iyi eleman yakalarsa o kadar genişler. 1 değerini alması ise tüm elemanların iyi olduğunu gösterdiğinden en mükemmel durum olarak görülür. ROC eğrisinin değerlendirme ölçütü olarak kullanılması için hangi elemanların iyi hangi elemanların kötü olduğu belirlenmelidir. Bu işlem için kullanıcının kendi oyları kullanılır. 4 ve 4'ün üstünde oy alan elemanlar iyi olarak nitelendirilirken altında oy alanlar kötü olarak kabul edilir (Sarwar B.M 2001).

b- PRC Duyarlılık Ölçütü: Filtreleme sisteminin kullanıcıya ne derecede ilgili bilgiler sunduğunun bir ölçütüdür. İşlemsel olarak *tamlık-çağırılan (precision-recall)* eğrisinin altındaki alandır. Tamlık seçilen elemanlardan ilgili olanların yüzdesi iken çağırılan ilgili elemanlardan seçilmiş olanların yüzdesidir. Bundan dolayı tamlık sistemin ne kadar seçici olduğunu gösterir. Çağırılan ise değerli bilgilerin bulunmasında ne kadar başarıya yaklaşıldığının göstergesidir. Bu ölçüt ne kadar yüksek değer alırsa doğruluk o kadar fazla demektir (Sarwar B.M 2001).

b) Tahminin Kapsamı :

Kapsam, öneri sisteminin tavsiye üretirken kullandığı elemanların yüzdesidir. Öneri sisteminin bilgi eksikliği (sparsity) veya çeşitli kısıtlamalar (eşik parametrelerine verilen değerler gibi) yüzünden belli bir eleman için tavsiye üretilmediği durumlarda önem kazanır. Kapsam değerinin düşük olması kullanıcının ya çok büyük miktardaki elemanlardan vazgeçmesi gerektiğini yada bu elemanları değerlendirmede başka kriterler kullanması gerektiğini gösterir. Kapsam değerinin yüksek olması ise öneri sisteminin pek çok eleman arasında seçim yapmada yardımcı olabileceğini gösterir. (Miller B. N. 2003)

Temel kapsam ölçütü, sistemin hakkında tahmin üretebilmesi mümkün olan elemanların yüzdesidir. Bu değer tam olarak hesaplanması güçtür çünkü her bir kullanıcı için onun oylarına ve komşularına bağlı olarak değişebilir. Bu yüzden

temel kapsam ölçütü yerine kullanım-merkezli kapsam ölçütü kullanılır. Bu ölçüt kullanıcının bir elemanı oylamadan önce tavsiye sisteminden faydalan yüzdesidir. Yani kullanıcı tarafından her bir elemana verilen oylar hakkında o eleman oylanmadan hemen önce sistem bir tahmin üretmiş ve kullanıcı bu tavsiyeden faydalanmışsa, bunu ne kadar sıklıkla yaptığının bir göstergesidir (Sarwar B.M 2001). Sonuç olarak bu oran ne kadar yüksek olursa öneri sisteminin o kadar verimli kullanıldığı ve güvenilirliğinin o kadar fazla olduğu söylenir .

2.1.2.3.2 Tavsiye Değerlendirme Ölçütleri

En iyi N tane tavsiyenin değerlendirilmesi işlemi doğası gereği tahmin değerlendirmekten farklıdır. Tahminler değerlendirilirken hesaplanan ve gerçek oy değerlerinin birbirlerine olan sayısal yakınlıkları hesaplanarak sonuca ulaşıyordu. Tavsiye değerlendirilirken odaklanılan esas düşünce ise sistemin tavsiye olarak ürettiği ve kullanıcının hoşlanacağını düşündüğü N tane üründen oluşan liste ile gerçekte kullanıcının satın aldığı/izlediği ürünlerin listesinin ne kadar uyduğunun tespit edilmesidir.

En iyi N tavsiye listesinin değerlendirilmesinde sıklıkla kullanılan üç ölçüt mevcuttur. Bunlar *tamlık(precision)*, *çağırılan(recall)* ve *F1* ölçütleridir. Hesaplanmaları için öncelikle veri kümesinin test ve eğitim olmak üzere ikiye ayrılmasını gerekir. En iyi N tavsiyenin elde edilmesi için tasarlanan algoritma eğitim kümesi üzerinde çalıştırılır. Algoritma geriye en iyi N elemanı barındıran bir N-tavsiye listesi döndürür. Bu liste ile test kümesindeki elemanlar karşılaştırılarak tavsiyenin doğruluğunu değerlendirilmesi için kullanılan ölçütler hesaplanır.

Çağırılan ölçütü, hem test kümesinde hem de N-tavsiye listesinde yer alan elemanların sayısının sadece test kümesinde yer alan elemanların sayısına oranıdır. *Tamlık ölçütü* ise hem test kümesinde hem de N-tavsiye listesinde yer alan elemanların sayısının sadece N-tavsiye listesinde yer alan elemanların sayısına oranıdır. Başka bir deyişle *çağırılan* kullanıcıların gerçekte izlediği/aldığı ürünlerin kaç tanesinin N-tavsiye listesinde yer aldığını gösterirken *tamlık* tavsiye edilen

ürünlerin kaç tanesinin gerçekten izlendiğini/satın alındığını gösterir. Bahsedilen bu iki ölçütün hesaplanması oldukça kolay olduğu için caziptir ancak N_{tavsiye} listesinin boyutu arttıkça bu iki ölçüt çakışır. Çünkü listenin büyümesi *çağırma* değerini artırırken *tamlık* değerinin azalmasına sebep olur. Bu yüzden bu ikisinin birleşiminden oluşan ve her iki ölçüte de eşit ağırlık verilmesini mümkün kılan “*F1 ölçütü*” tanımlanmıştır. Pek çok çalışmada kaliteyi değerlendirme ölçütü olarak kullanılan $F1$ 'in hesaplanması ile diğer iki ölçütün hesaplanması çizelge 2.2’de gösterildiği gibidir (Cho ve Kim 2004).

Çizelge 2.2 : Tavsiye Değerlendirme Ölçütleri

<i>Ölçüt</i>	<i>Hesaplanması</i>
<i>Çağırılan(Recal)</i>	$\frac{ test \cap N_{\text{tavsiye}} }{ test }$
<i>Tamlık(Precision)</i>	$\frac{ test \cap N_{\text{tavsiye}} }{N}$
<i>F1</i>	$\frac{2 * cagirilan * tamlık}{cagirilan + tamlık}$

2.1.2.3.3 Öneri Sisteminin Performansının Değerlendirilmesi

Sistemin ürettiği tahmin ve tavsiyelerin kalitesinin değerlendirilmesinden başka sisteminin ne kadar verimli çalıştığının analiz edilebilmesi için performansının da değerlendirilmesi gereklidir. Performans değerlendirmesi için kullanılan en yaygın yöntem “Cevap Zamanı (Response Time)” dır. Cevap zamanı eğitim kümesi için tüm tavsiyelerin hesaplanmasında ihtiyaç duyulan zamandır. Performansı değerlendirmede kullanılan diğer bir yöntem ise saniyede yapılan tavsiye sayısını gösteren “Tavsiye Hızı (Throughput)” dır (Cho ve Kim 2004) .

2.1.3 Karşılaşılan Problemler ve Sunulan Çözümler

Tavsiye sistemlerinin pek çok başarılı uygulaması olmasına rağmen halen çözümlenmesi gereken ve akademik olarak araştırılmaya müsait sorunları

bulunmaktadır. Bunlardan başlıcaları bilgi eksikliği (sparsity) ve ölçeklenebilirlik (scalability) ve problemleridir. Bu problemler ve sunulan çözümler bir sonraki kısımda açıklanmıştır.

2.1.3.1 Bilgi Eksikliği

İFT tabanlı öneri sistemleri ağırlıklı olarak benzer kullanıcıların komşuluklarını kullanarak işlem yaparlar. Kullanıcı benzerlikleri ise geçmişte farklı elemanlara verdikleri oylara bakılarak tespit edilir. Fakat her kullanıcının her bir elemanı oylaması oldukça nadir görülen bir olaydır. Mesela amazon.com gibi çok geniş bir ziyaretçi/müşteri kitlesine sahip bir elektronik ticaret sitesinde 2 milyon kitap seçeneğini olduğunu varsayalım. Gelen bir müşterinin ise 2000 kitap oyladığı (incelediği) varsayılabilir bile seçeneklerin sadece binde birini oylamış demektir. Buda halen pek çok kitabın oylanmadığını gösterir. Aynı düşük oran kullanıcıların oyladıkları ortak elemanlar içinde geçerli olacaktır. Bu da benzerlik hesabını olumsuz yönde etkileyecek ve komşulukların bulunmasını güçleştirecektir (Sarwar 2001). Bu sorun bilgi eksikliği olarak adlandırılır.

Bu problemi çözmek için geliştirilen yaklaşımlar üç ana başlık altında toplanabilir: Dâhili Oylamalar (Implicit Ratings), Karma Filtreleme (Hybrid Filtering), Ürün-Ürün Bağlantısı (Product-Product Correlation). Bu yöntemleri kısaca açıklayalım.

Dahili Oylamalar müşterilerin davranışlarından faydalanılarak, veri tabanındaki oyların sayısını artırmaya çalışır. Karma Filtreleme ise işbirlikçi filtreleme ile içerik tabanlı filtreleme tekniklerinin birleşimini kullanmaktadır. Böylece işbirlikçi filtrelemenin benzerlik bulurken ve tahmin yaparken kullanacağı oyların sayısı oylanmış filmlerin içeriklerini, kullanıcı demografiklerini kullanan filtreleme teknikleri yardımı ile artırılmaya çalışılır. Ürün-Ürün Bağlantısında ise yukarıdaki yaklaşımlar yerine, müşterilerin farklı ürünler ile olan ilişkilerini tanımlayabilmek için müşteri profilini analiz eder ve bu ilişkileri verilen bir müşteri-ürün çifti için tahmin değeri hesaplama kullanır (Cho ve Kim 2004).

2.1.3.2 Ölçeklenebilirlik

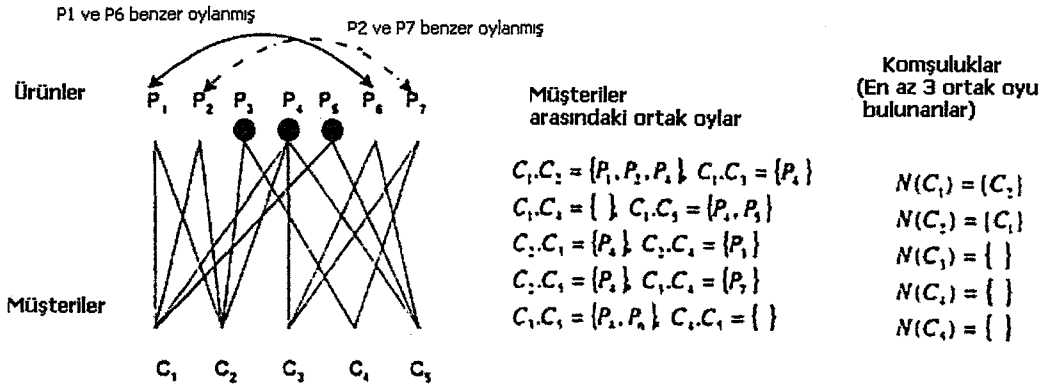
Öneri sistemleri, genellikle büyük ölçekli veri tabanları için geçerlidir. Bu da algoritmaların çalışma zamanlarını uzatırken performanslarının düşmesine sebep olur. Bu sorunun üstesinden gelinmesi için sunulan yaklaşımlar iki kategori altında toplanabilir. Bunlar “Boyut Azaltma Teknikleri (Dimensionality Reduction Techniques)” ve “Model Tabanlı Yaklaşımlar (Model-Based Approach)”dır.

Boyut azaltma tekniği olarak “Tekil Değer Azaltma (TDA) (Singular Value Decomposition)” tekniği kullanılır. Bu bir matris faktörizasyon tekniğidir. Genelde düşük-rank yakınsamaları elde etmek için kullanılır. Rankı r olan bir $m \times n$ boyutlu bir A matrisi verildiğinde, $TDA(A)$ aşağıdaki gibi tanımlanır.

$$TDA(A) = U \times S \times V^T$$

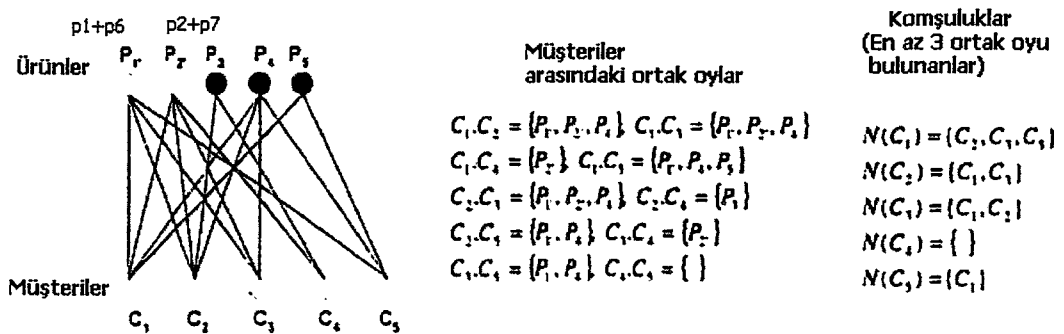
U , S ve V sırasıyla $m \times m$, $m \times n$ ve $n \times n$ boyutlu matrisleri temsil eder. Buradaki S matrisi diagonal bir matristir, r tane sıfıra eşit olmayan girişe sahiptir ve tekil (singular) matris olarak adlandırılır. S matrisinin girişleri $\{s_1, s_2, \dots, s_r\}$ şeklinde gösterilsin. Burada her bir $s_i > 0$ ve $s_1 \geq s_2 \geq \dots \geq s_r$ 'dir. U ve V matrisleri de düşük rank değerine yakınsanmış ortogonal matrislerdir.

Araştırmacılar orijinal matris yerine düşük-ranklı matrislerin kullanılmasının daha avantajlı olduğu görüşündedirler (Sarwar 2000). TDA yaklaşımı sayesinde birbirinden bağımsız öz vektörler üretir. Böylece her bir kullanıcı ve ürün kendi öz vektöründe gösterilebilir. Boyut azaltma işleminin sonucunda da benzer ürünleri oylamış müşteriler aynı öz vektör içinde yer alabilir. Böylece veri tabanının büyüklüğü sebebiyle birbirleri arasındaki komşuluğu gizli kalmış kullanıcıların açığa çıkması sağlanmış olur.



Şekil 2.6 : Boyut azaltmanın yararını belirtmek için tasarlanmış örnek yapı

Boyut azaltmanın ölçeklenebilirlik probleminin çözümüne nasıl bir katkıda bulunduğu bir örnek ile açıklanmıştır. Şekil 2.6'da gösterildiği gibi bir veri yapımız olsun. Bu örnekte yedi tane ürün 5 farklı kullanıcı tarafından oylansın. Şeklin solundaki grafa müşterilerin hangi filmleri oyladığı gösterilmiştir. p1 ve p6 ile p2 ve p7 ürünleri kendi aralarında benzer olarak oylanmış ürünler olarak kabul edilmiştir. Şeklin ortasında ise müşterilerin oyladıkları ortak ürünler gösterilmiştir. Mesela C1 ile C2 p1, p2 ve p4 ürünlerini oylamış görünürken C4 ile C5 arasında hiçbir bağ bulunmamaktadır. Komşuluk belirleme kriteri olarak da en az üç tane ortak oyladıkları film olması kabul edilsin. Bu durumda sadece C1 ve C2 arasında bir komşuluk olduğu söylenebilir.



Şekil 2.7 : Boyut azaltma uygulandıktan sonra tasarlanmış örnek yapının hali

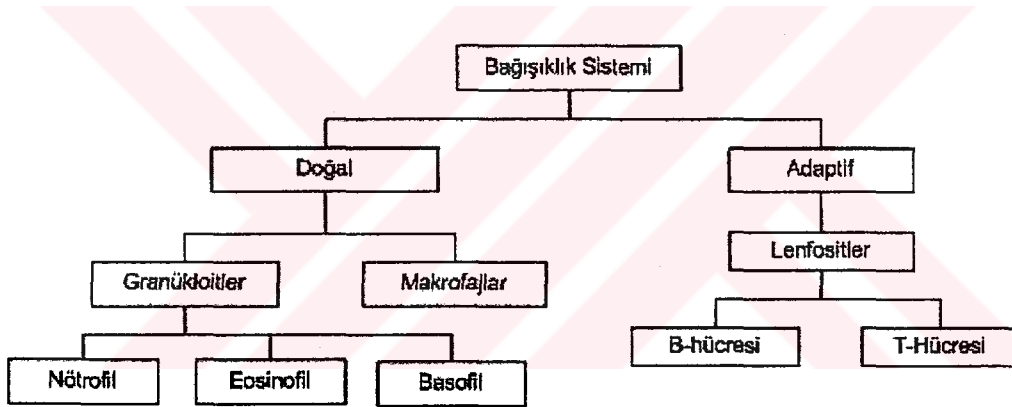
Bu kümeye boyut azaltma tekniğinin uygulandığını ve şekil 2.7’de görülen grafin elde edildiğini varsayalım. Bunun sonucunda p1 ile p6 ve p2 ile p7 çiftleri arasındaki gizli ilişkiler ortaya çıkarılmış, daha önce yedi olan ürün sayısı beşe indirilerek gizli kalmış komşulukların ortaya çıkarılması sağlanmıştır. Bir önceki örnekte iki kullanıcı için komşu bulunabilmişken bu örnek için sadece bir kullanıcı için komşuluk oluşturulamamıştır.

Ölçeklenebilirliği sağlamak için sunulan diğer bir yöntem ise oylama matrisine dayalı bir model oluşturulması ve bu model kullanılarak tavsiye yapılmasıdır. Bu yöntemin dezavantajı modelin inşasının pahalı bir işlem olmasıdır ancak çalıştırılması hızlıdır. Bayes ağları, kümeleme ve birliktelik kural madenciliği gibi teknikler model inşasında kullanılabilir.



3. YAPAY BAĞIŞIKLIK SİSTEMLERİ

İnsan bağışıklık sistemi virüs, bakteri gibi neredeyse sınırsız sayıda olan yabancı saldırganlara karşı doğal, hızlı ve etkili cevap verme yeteneğine sahip karmaşık bir ağ yapısı olarak tanımlanabilir. Bağışıklık sistemi örüntü tanıma, öğrenme, hafıza edinme, farklılıkların genelleştirilmesi, gürültü toleransı, genelleme ve optimizasyon gibi pek çok işlemi yerine getirme kabiliyetine sahip hücrelerden meydana gelir. Bağışıklık sistemini daha iyi anlamak ve mühendislik problemlerinin çözümünde kullanabilmek için bu biyolojik prensiplere dayanan yeni hesaplama teknikleri geliştirilmektedir. Bu amaçla insan bağışıklık sisteminden esinlenerek yapay bağışıklık sistemleri tanımlanmıştır. Bu bölümün diğer konularında ilk olarak bağışıklık sisteminin yapısı, temel bileşenleri ve özelliklerinden bahsedilmiştir. Daha sonra yapay bağışıklık sistemi tanıtılmış ve algoritmaları anlatılmıştır.



Şekil 3.1 : Bağışıklık Sisteminin Sınıflandırılması ve Temel elemanları (de Castro, ve Timmis, 2002)

3.1 İnsan Bağışıklık Sistemi

3.1.1 İnsan Bağışıklık Sisteminin Yapısı

İnsan bağışıklık sistemi, dışarıdan gelebilecek saldırılara karşı vücudu koruma yeteneğine sahip kuvvetli bir savunma mekanizmasıdır. Bu mekanizmanın esas öğeleri beyaz kan hücreleri yani akyuvarlardır (leukocyte). Bağışıklık sistemi

doğal ve adaptif olmak üzere ikiye ayrılır. Şekil 3.1’de de görüldüğü gibi doğal bağışıklık sisteminin elemanları granüloitler (granulocytes) ve makrofajlar (macrophages) iken adaptif bağışıklık sisteminin elemanları lenfositlerdir (lymphocytes).

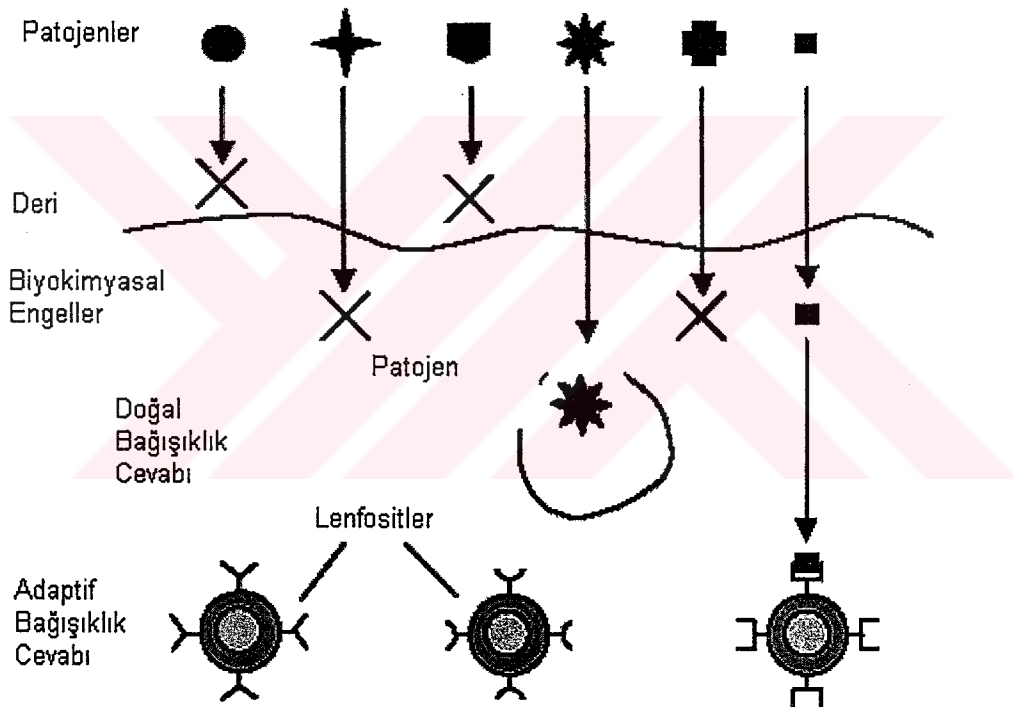
İnsan vücudunun doğumla birlikte kazandığı bağışıklık, doğal bağışıklık olarak adlandırılır. İnsan vücudunun sahip olduğu bu doğuştan gelen bağışıklık pek çok mikrobu ilk karşılaşmada tanır ve yok eder. Çalışma şekli iki moleküler yapının birbirini tanıyarak tamamlamasına dayanır. Patojen üzerinde bulunan yapı PAMP (pattern associated molecular pattern) olarak adlandırılır. PAMP’ları tanıyan ve onun tamamlayıcısı olarak kodlanmış vücutta üretilen yapılara ise PRR (pattern recognition receptors) adı verilir. PAMP’lar sadece mikroplar tarafından üretilir. Asla vücutta üretilmez. O yüzden bir PRR, bir PAMP’ı tanıdığına dair bir sinyal üretirse bu vücutta mikrop girdiği anlamına gelir. Doğal bağışıklık sisteminin önemli bir elemanı da tanınamayan yapıları T-hücrelerine sunan APC (antigen presenting cell)dir. Böylece bu hücreler adaptif bağışıklık cevabını başlatırlar.

Tespit edilmiş bir antijeni yok etmekle görevli antikorların üretilmesine Adaptif bağışıklık cevabı denilir. Adaptif bağışıklığın elemanları B ve T olarak adlandırılan iki çeşit lenfosit hücredir. Bu hücreler üzerlerinde bulunan reseptörler ile daha önceden hiç karşılaşmadıkları antijenleri tanıyabilirler. Diğer bir özellikleri ise tanıdıkları bu antijenler için bağışıklık hafızası geliştirme özelliğine sahip olmasıdır. Böylece vücutta daha önceden tanınmış bir antijen tekrar gelirse aktivasyon süresinin kısaltılması sağlanmış olunur. Bu hücreler ile ilgili daha detaylı açıklamalar ilerleyen bölümlerde yapılmıştır.

3.1.2 İnsan Bağışıklık Sisteminin Mimarisi

Bağışıklık sisteminin düzenleme ve koruma mekanizmalarının çeşitli seviyelerde yer aldığı çok katmanlı bir mimarisi olduğu söylenebilir. Bu katmanlar şekil 3.2’de gösterilmiştir. Sırası ile bu katmanların açıklamaları aşağıdaki gibidir (de Castro, ve Timmis, 2002).

- **Fiziksel Engeller:** Derimiz zararlı olsun yada olmasın yabancı maddelere karşı vücudumuzu korumak için kalkan gibi çalışır. Solunum sistemimizin de dış yüzey ile temasta olan organlarının mukoza kaplı kısımları sayesinde dışardan gelecek muhtemel tehlikelere karşı vücut savunmasına yardımcı olur.
- **Biyokimyasal Engeller :** Vücudun salgıladığı tükürük, ter, göz yaşı gibi sıvılar yıkıcı enzimler içerirler. Mide asitleri ise yeme ve içme ile vücuda giren mikroorganizmaları öldürürler. Vücudun sıcaklığı ve pH değeri bazı saldırganlar için elverişsiz hayat şartları sunar.



Şekil 3.2 : Bağışıklık Sisteminin Çok Katmanlı Mimarisi

- **Doğal ve adaptif bağışıklık sistemi:** Bağışıklık sistemini oluşturan iki önemli kavramdır ve bir önceki bölümde kısaca açıklanmıştır.

3.1.3 Baęışıklık Hücreleri

Baęışıklık sistemini meydana getiren lenfoit organlar vücudun her yerine dağılmıştır. Bunlardan başlıcaları bademcik, lenf bezi, lenf damarları, kemik ilięi, lenf düęümleri, timüs ve dalaktır. Baęışıklık hücreleri kemik ilięinde üretilir, lenf düęümlerinde timüste depolanır, lenf damarları sayesinde dokulara giderler. Üretilen hücrelerden bir kısmı vücudun genel savunmasından sorumlu iken dięerleri özel patojenlerle savaşmak için eğitilir. Etkili bir şekilde çalışmalarını için hücreler arasında devamlı bir işbirlięi gereklidir. Bu bölümde lenfositler, fagositler ve granüositlerden bahsedilecektir.

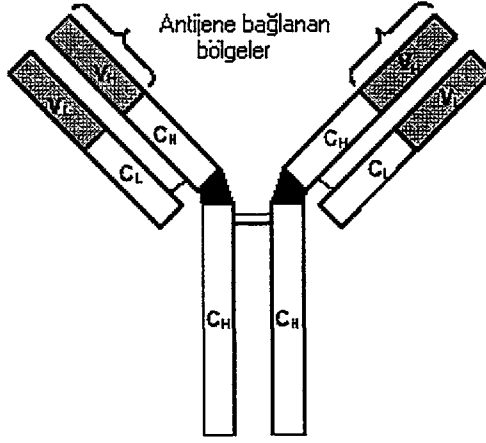
3.1.3.1 Lenfositler

Lenfositler baęışıklık sisteminde önemli bir sorumluluęa sahip olan küçük lökositlerdir. Lenfositlerin iki ana tipi vardır. Bunlar B lenfositleri (veya B hücresi) ve T lenfositleridir (veya T hücresi). Lenfositlerin çoęu sadece özel bir aktivasyonun çoęalması için gerekli ayrı antijenlerle etkileşime girdikten sonra fonksiyonel aktivite sergileyen küçük dinlendirici hücreler şeklinde düzenlenir. B ve T lenfositlerinin yüzeylerinde belirli antijenik yapılar için oldukça özelleşmiş reseptörler bulunur. Bölümün devamında sırası ile bu hücrelerden daha ayrıntılı bir şekilde bahsedilmiştir.

3.1.3.1.1 B hücreleri ve antikorlar

B hücrelerinin temel fonksiyonu bakteri, virüs ve tümör hücreleri gibi dış kaynaklı proteinlere tepki vermek için antikor üretmek ve salgılamaktır. Her B hücresi belirli bir antikor üretmek için programlanmıştır. Antikorlar bir başka belirli proteini tanıyan ve ona bağlanan belirli proteinlerdir. Antikor üretimi ve bağlanması genellikle maddeyi kuşatıp sindirme veya yok etme, dięer hücreleri öldürmek için sinyal verme şeklinde olur.

B hücreleri, yüzeylerinde Immunoglobulin molekülleri taşırlar (De Castro and Von Zuben, 1999a). Immunoglobulin molekülleri birbirine benzer yapıda antikorlardan oluşur . Şekil 3.3'de görüldüğü gibi Antikorlar, protein dizilimlerinin yer



Şekil 3.3 Antikor'un Yapısı

Bir antikorun (Ab) temel ünitesi veya immunoglobulin (Ig) molekülü iki özdeş ışık (Light L) zincirinden ve iki özdeş ağırlık (Heavy H) zincirinden meydana getirilmiştir Şekil 3.3'de bir antikorun şekli görünmektedir. aldığı bir ağır zincir (Heavy Chain) bir de hafif zincir (Light Chain)'den oluşurlar. Antikorların V (Variable) bölgeleri, antijenlerle ya da diğer antijenik yapılarla birleşme yerleridir (reseptörleri). Bir birey yaklaşık 10^9 farklı çeşitte Antikor üretebilir. Antikorların efektör fonksiyonlarının çoğu ağır zincirlerinin C bölgeleri ile başlar. Efektör fonksiyonlarının başlayabilmesi için de V bölgelerinin Antijen reseptörlerine bağlanmaları gerekmektedir. B hücreleri aktive olduklarında plazma hücrelerine dönüşürler ve yüzeylerinde bulunan Immunoglobulin moleküllerinden antikor salgırlar. Salgıladıkları bu antikorlar antijenler ile birleşerek bir antijen-antikor kompleksi oluştururlar. Bu kompleksin oluşması, çevreye bazı kimyasal sinyaller yayılmasını indükler ve bu sinyalleri algılayan öldürücü hücreler bu kompleksi parçalayarak yok eder (De Castro and Von Zuben, 1999) .

3.1.3.1.2 T hücreleri

Timüs içerisinde olgunlaşan T hücrelerinin görevi hücrelerin hareketlerini düzenlemek ve direkt olarak vücuda girmiş mikroplara saldırmaktır. T lenfositleri üç temel alt sınıfa ayrılabilirler. Bunlar T hepler hücreleri (Th), sitotoksik (cytotoxic-killer) T hücreleri ve bastırıcı (suppressor) T hücreleridir.

Aynı zamanda CD4 veya T4 hücreleri olarak da bilinen T hepler hücreleri veya basit olarak Th hücreleri diğer bağışıklık hücrelerinin yani B hücrelerinin, T hücrelerinin geri kalan türlerinin, makrofajların ve doğal killer (NK) hücrelerinin aktivasyonu için gereklidir. Killer T hücrelerinin veya cytotoxic T hücrelerinin mikrobik saldırganları, virüsleri veya kanserli hücreleri yok etme yeteneği vardır. Aktif hale geldikten sonra yabancı hücreleri kuşatma, onların yüzey zarına nüfuz edip yıkımlarına sebep olma ve çıkan zararlı kimyasal maddeleri sindirme özellikleri vardır. CD8 hücreleri olarak da adlandırılan Bastırıcı (supressor) T lenfositleri bağışıklık tepkisinin devamını sağlamada önemli bir rol üstlenirler. Esas görevleri bağışıklık hücrelerinin kontrolsüz çoğalmasını engellemek yani onlara baskılayıcı bir etki uygulamaktır. Böylece bağışıklık, alerjik reaksiyonlar ve otoimmün hastalıklar kontrol altında tutulabilir (De Castro and Von Zuben, 1999).

3.1.3.2 Fagositler, Granülositler ve onların ilişkileri

Fagositler ("hücre yiyenler") mikro organizmaları ve antijenik tanecikleri yeme ve sindirme yeteneği olan beyaz kan hücreleridir. Bazı fagositlerin aynı zamanda lenfositler için antijen sunma yeteneği vardır. Bu fagositler antijen sunan hücreler (APCs) olarak adlandırılmaktadır. Önemli fagositler monositler ve makrofajlardır. Monositler kan içinde dolaşırlar ve makrofaj ("büyük yiyiciler") haline geldikten sonra dokulara geçerler. Makrofajlar çeşitli fonksiyonlar olan çok yönlü hücrelerdir. Onlar T lenfositleri için antijen sağlarlar, onları yerler ve sindirirler. Bağışıklık tepkisinin başlamasında da önemli bir role sahiptirler.

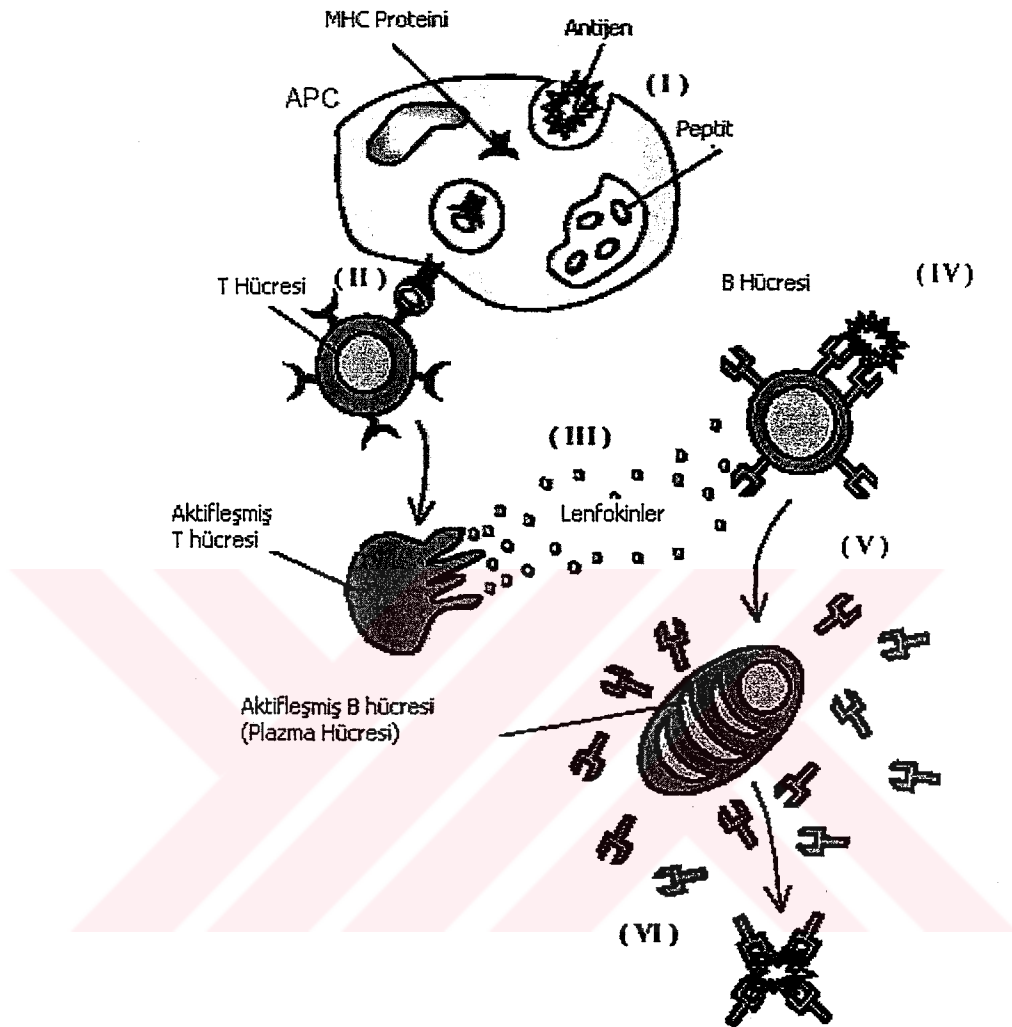
Nötropiller ve ösonopiller aynı zamanda makrofajlara benzer fonksiyonları olan granüositlerdir. Diğer bir granüosit olan basopiller ise kan dolaşımında yer alırlar, güçlü kimyasal maddelerle doldurulmuş tanecikler içerirler. Alerjik durumlara yanıt vermede önemlidir. İçlerinde yer alan kimyasal maddeler mikro organizmaları yok etmede görev alır (De Castro and Von Zuben, 1999).

3.1.4 Bağışıklık Sistemi Vücudu Nasıl Korur ?

Bağışıklık sistemi vücudu dışardan gelen saldırılara karşı koruyan hücre ve moleküllerin karışık bir kümesinden oluşur. Vücudumuz bütün bağışıklık tepkilerinin nihai hedefi olan antijenin (Ag) tehdidine karşı adaptif bağışıklık sistemi tarafından korunmaktadır.. Şekil 3.4' de bu savunma mekanizmasının basitleştirilmiş gösterimi bulunmaktadır. Şekilde görülen bu adımların kısaca açıklaması aşağıda verildiği gibidir.

Makrofajlar gibi antijen sunma yeteneğine sahip APC hücrelerinin yiyebilecekleri ve sindirebilecekleri antijenleri bulmak için vücuda dolaştığını ve bulduklarında ise onları *antijenik peptitlerine* ayırarak parçaladıklarını göstermektedir.(I) Bu peptitler MHC (major histocompatibility complex) moleküllerine bağlanır, APC hücre yüzeyi ile birleşerek MHC/Peptit kombinasyonunu oluştururlar.(II) T hücreleri, farklı bir MHC/Peptit kombinasyonlarını tanımak için her biri aktif olan reseptör moleküllerine sahiptirler. (III). T hücresi MHC/peptit kombinasyonu tanıdıktan sonra aktive olur ve parçalanarak lenfokinler ve ya kimyasal sinyaller salgılar. Bunlar diğer bağışıklık sistemi elemanlarını uyarır ve aktivasyona katılmasını sağlar. (IV)T hücrelerinin tersine B lenfositleri MHC moleküllerinin yardımı olmaksızın serbest durumdaki antijenleri tanıma yeteneğine sahip reseptörlere sahiptir.(V) B hücrelerinin reseptörleri tarafından uyarı sinyali alındığında B hücreleri *plazma hücrelerine* dönüşerek çok sayıda antikör molekülü salgırlar. (VI) Serbest kalan bu antikörler patojenleri etkisiz hale getirmede kullanılır. (VII) ve daha sonra bunlar enzimler veya temizleme hücreleri kullanılarak yok edilirler.Aktive olan bu b ve t hücrelerinin bir kısmı değişerek *hafıza hücrelerini* oluştururlar. Bu hücreler organizma içinde

dolaşmaya devam edecek ve gelecekte olabilecek aynı veya benzer antijene daha iyi bir bağışıklık cevabı üretmeyi garantileyeceklerdir (de Castro ve Timmis 2002).



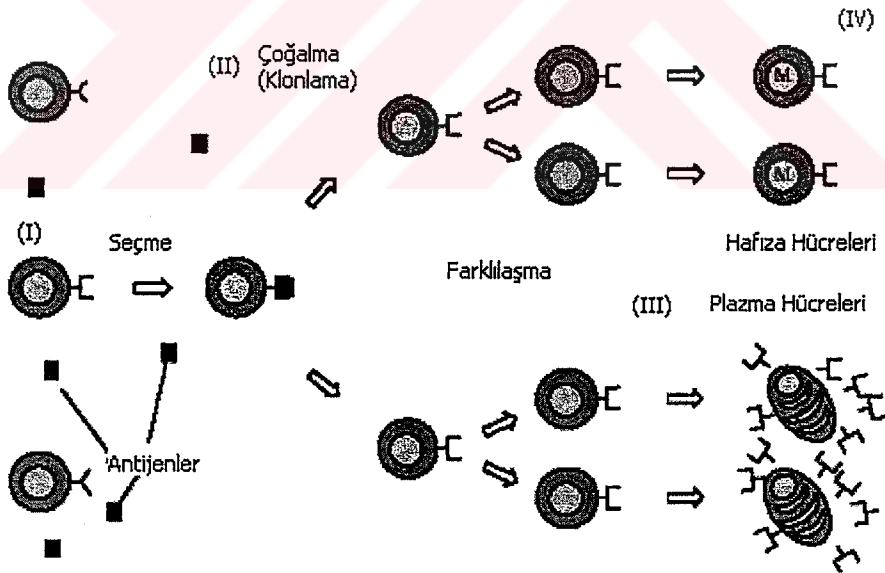
Şekil 3.4 Bağışıklık Sisteminin basit aktivasyon mekanizması

3.1.5 Klonal Seçme Teorisi

Klonal seçme teorisi antijenik bir uyarıcıya karşı adaptif bağışıklık cevabının temel özelliklerini açıklamak için kullanılan bir teoridir. Bu düşünce sadece antijenik uyarıyı fark edip olgunlaşarak efektör hücelere değişen hücreler için kurulmuştur. Klonal seçme prensibi hem T hücreleri hem de B hücrelerine uygulanır. Ancak B ve T hücrelerinin klonal çoğalması sırasındaki temel fark, B hücrelerinin çoğalma

boyunca somatik mutasyondan etkilenmesi buna karşın T hücrelerinin etkilenmemesidir.

Bir hayvan antijene maruz kaldığı zaman, B lenfosit hücreleri antikor üretmek tepki verir. Her hücre antijenin özelliğine bağlı olarak antikorun sadece bir türünü salgılar. Şekil 3.5’de klonal seçme prensibinin basitçe gösterimi bulunmaktadır. B hücresi üzerindeki antijenik reseptörler bir antijen ile bağlanır ve T_H gibi yardımcı bir hücrenin aracılığı ile ikinci bir sinyal üretir. Bu sinyal B hücresinin o antijen için uyarılmasını sağlar. Bu uyarılma sonucu B hücresi bölünür (II) ve olgunlaşarak antikor üreten plazma hücrelerine dönüşür.(III) Plazma hücreleri hızlı antikor salgırlarken, düşük bir oranla da olsa hızla bölünen B hücreleri de antikor salgırlar.B hücreleri plazma hücrelerinden başka uzun ömürlü olan hafıza hücrelerine de dönüşebilirler. (IV) Hafıza hücreleri kan aracılığıyla vücuda dolaşır ve antikor üretmezler. Ancak ikinci bir antijen uyarısı ortaya çıkarsa, hızla plazma hücrelerine dönüşürler ve yüksek duyarlılıklı antikor üretmeye başlarlar. (de Castro, ve Timmis, 2002)

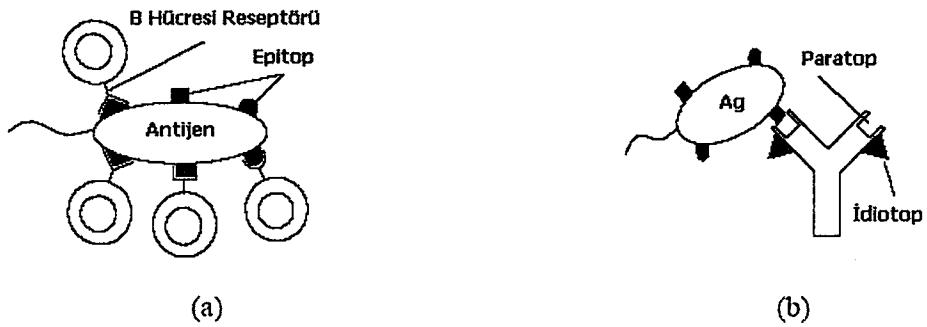


Şekil 3.5 Klonal Seçme prensibi

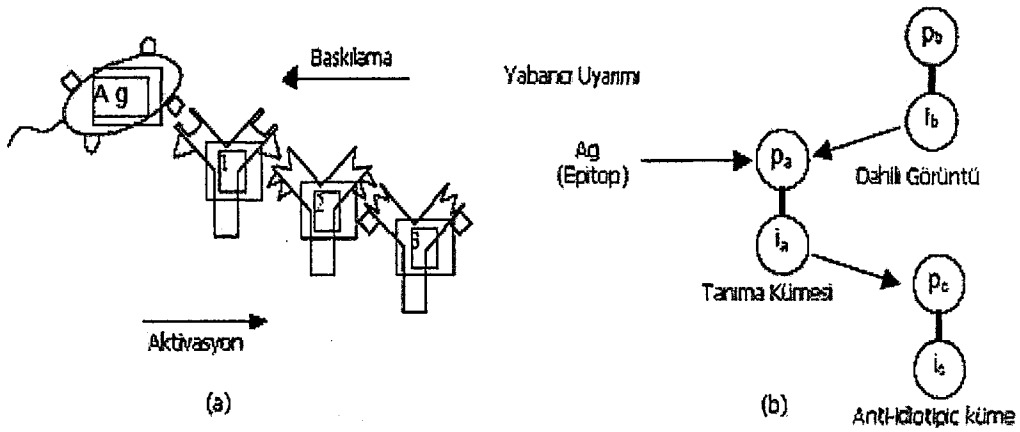
3.1.6 Baęışıklık Aę Teorisi

Baęışıklık aę teorisi Jerne tarafından 1974 yılında ilk defa önerildięinde ne hücre sinyallerinden, ne antikor moleküllerinden ne de hücreler arasında etkileşim mekanizmaları ile bu etkileşimin bir sonucu olarak etkili olabilen mekanizmalardan bahsetme amacı yoktu. Onun yerine lenfosit aktivitelerini, doęal antikor üretimini, repertuar seçimini, tolerans ve kendi/kendi-deęil ayırımını, baęışıklık sisteminin hafızasını ve evriminin bakış açısını hipotezlemeyi amaçlamıştır. Jerne'e göre baęışıklık sistemi antijenlerin yokluęunda bile birbirini tanıyabilen hücrelerin düzenlenmiş bir aęından meydana gelmiştir.

Baęışıklık aęını tarif etmek için Jerne'in kullanılmasını önerdiği gösterim Şekil 3.6'da gösterildięi gibidir. Antijen üzerinde bulunan ve Antikorum reseptörleri sayesinde tanıyıp antijene baęlandıkları kısım *epitop* olarak adlandırılır. Bir antikor molekülün tamamlayıcı epitopunu tanımakla görevli deęişken bölgesinde yer alan kısmına *paratop* adı verilir. Deęişken bölgede olduęu için bir antikorum paratop'un şekli deęişebilir. İdiotop ise antikorum paratobunun çevresinde veya içerisinde yer alan antikoru tanımlayan kısma verilen addır. Bir antijen birden fazla epitopa sahip olabilirken antikorlar tek tip epitoba sahiptir ve burası idiotop olarak isimlendirilir..



Şekil 3.6 : Baęışıklık Aę teorisinin basit kavramları. (a) Farklı B hücreleri tarafından tanınan çeşitli epitoplara sahip bir antijendir. (b) Antikorum V bölgesi ile paratop ve idiotopun gösterimidir.



Şekil 3.7 : Bağışıklık Ağ teorisinin gösterimi. (a) Bir antijenin 1 nolu sınıfın bir antikorunu uyarması, bu antikorun 2 nolu sınıfın antikorunu uyarması vb. (b) İdiyotipik ağ yapısının ayrıntılı gösterimi

Bağışıklık sistemi biçimsel olarak idiotopların kümelerini tanıyan paratopların ve paratop kümeleri tarafından tanınan idiotopların büyük ve karmaşık bir ağı olarak tanımlanabilir. Böylece her eleman tanıyabilmesinin yanı sıra tanıya da bilir. Bu özellik ağın kurulmasında önderlik eder. Verilen antikor bir epitopu veya idiotopu tanıdıktan sonra, üretilen tanıma sinyaline pozitif veya negatif olarak cevap verebilir. Pozitif cevap hücre aktivasyonu, hücre çoğalması ve antikor salgılanması ile sonuçlanırken, negatif cevap baskılama oluşmasına sebebiyet verir. Şekil 3.7 (a)'da bağışıklık modelinin pozitif ve negatif cevabı görülür.

Şekil 3.7 (b)' de ise bağışıklık ağ teorisinin davranışının gösterilmiştir. Bağışıklık sistemi bir Ag tarafından uyarıldığında, bu antijenin epitopu farklı paratoplar tarafından farklı derecelerde tanınır. Bu paratopların oluşturduğu sete p_a adını veriyoruz. Bu p_a paratopları, Ab ve reseptör moleküllerinin paratoplarıdır. Bu setteki her paratopun tanıdığı idiotoplar vardır. Böylelikle p_a paratop setinin tanıdığı bir idiotop seti yani i_a seti ortaya çıkar. $p_a i_a$ sembolü, Ag'i tanıyan Ab molekülleri ile, potansiyel olarak ona yanıt oluşturabilecek lenfositlerin oluşturduğu toplam bir settir. Bağışıklık ağında p_a setindeki her paratop, bir idiotop setini tanır. Böylece

toplam p_a seti oldukça geniş bir idiotope setini tanımış olur. Buradaki i_b idiotope seti, Ag antijenini tanıyan p_a seti tarafından tanınan (eğer varsa) diğer epitoplara imajını (internal image) temsil eder. Bu idiotope setini – yani i_b 'yi tanıyan diğer paratopların oluşturduğu set ise p_b setidir. Bunun yanında, p_a 'a setindeki her idiotope, bir paratop seti tarafından tanınır. Bu da toplam i_a setinin daha büyük bir p_c paratop seti tarafından tanınması anlamına gelir. Yine benzer şekilde, bu p_c setindeki her bir paratop bir i_c idiotope seti tarafından tanınır ve toplam p_c seti daha geniş bir i_c seti ile birlikte ifade edilir- $p_c i_c$. Fakat bu sefer bu sete tanıyıcı set yerine anti-idiotypik set adı verilir çünkü bu set asıl Ag'i değil, idiotop'ları tanımaktadır. Bu yöntemi bu şekilde devam ettirirsek, ağ içerisinde bir önce tanımlanan setleri tanıyan ve yine bu setler tarafından tanınan oldukça büyük setlere ulaşırız. (de Castro ve Von Zuben 2001)

Ağ yaklaşımı özellikle bilgisayar araçlarını geliştirmek için caziptir çünkü öğrenme ve bellek, hücre popülasyonunun boyutu ve çeşitliği gibi gelişen özelliklerin bir tanımını sağlar. Bağışıklık ağ modeli temel alınarak modeller vardır. Bu modellerin pek çoğunun yapısını aşağıdaki terimleri kullanarak gösterebiliriz:

$$\text{PDO} = \frac{\text{Akımı}}{\text{Ölümü}} - \frac{\text{uyarılmamış hücrelerin}}{\text{Ölümü}} + \frac{\text{uyarılmış hücrelerin}}{\text{Üretimi}}$$

Burada PDO popülasyon değişim oranıdır. (de Castro ve Von Zuben 2001)

3.2 Yapay Bağışıklık Sistemleri

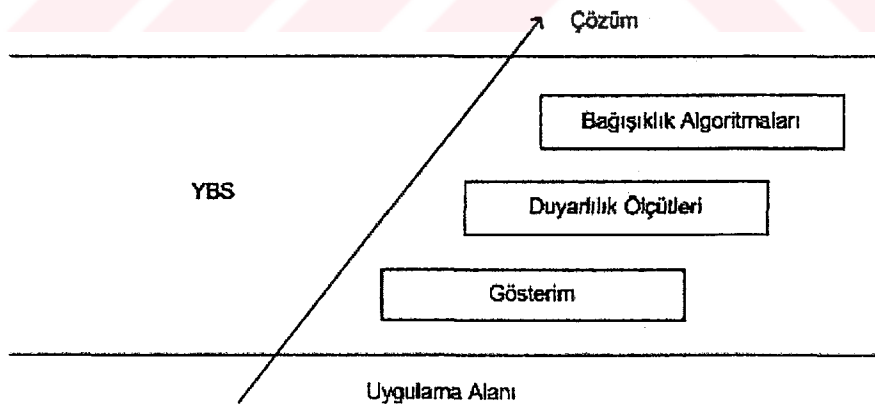
Son yıllarda biyolojiden esinlenerek geliştirilmiş sistemler hakkında yapılan çalışmalarda büyük bir miktarda artış gözlemlenmiştir. Bunların arasında yapay sinir ağları (ANN), evrimsel hesaplama, DNA hesaplama ve yapay bağışıklık sistemleri sayılabilir. Yapay bağışıklık sistemleri problemleri çözmek için kullanılan adaptif sistemlerdir. Bağışıklık fonksiyonlarını, prensiplerini ve modellerini gözlemlemiş ve teorik bağışıklıktan esinlenmişlerdir (de Castro ve Timmis 2002).

Bağışıklık sistemi uygulamalarının başlıca kullanım alanları örüntü tanıma, hata ve anormallik tespiti, veri analizleri (veri madenciliği, sınıflandırma vb.) etmen tabanlı sistemler, makine öğrenmesi, arama ve optimizasyon problemleri ve güvenlik ve bilgi sistemleridir.

Biyolojiden esinlenen diğer algoritmalarda olduğu gibi yapay bağışıklık sistemlerinde de dizayn aşamasında aşağıdaki temel soruların cevaplarını vermesi gerekmektedir.

- Sistemdeki elemanlar nasıl temsil edilecektir?
- Bireylerin çevre ve birbirleri ile olan ilişkilerini değerlendirmek için hangi değerlendirme mekanizmaları kullanılmak için seçilmiştir?
- Sistem dinamiğini yönetmek için adaptasyon prosedürlerinin belirlenmesi. Yani sistem davranışı zamanla neye göre değişecektir?

Yukarıdaki soruların cevaplanma süreçleri yapay bağışıklık sistemi oluşturulurken izlenmesi gereken katmanlı bir yapının aşamaları olarak da düşünülebilir. Şekil 3.8’de YBS için katmanlı yapı yaklaşımı görülmektedir.



Şekil 3.8 YBS' nin Katmanlı Yapısı

Bu yapıya göre tasarım bağışıklık hücre ve moleküllerin gösteriminde kullanılacak soyut modelin belirlenmesi ile başlar. 1979 yılında Perelson ve Oster'in sunduğu *Şekil Uzayı* en çok kabul gören ve en sık kullanılan yaklaşımdır. İkinci

adım ise YBS elemanları arasındaki ilişkileri değerlendirmek için literatürde sunulan çeşitli ölçütlerden uygun olanını seçmektir. Mesela iki elemanda ikili kodlanmış ise aralarındaki ilişki Hamming uzaklığı kullanılarak değerlendirilebilir. En son adımda ise sistemdeki dinamiği yönetmek için adaptasyon prosedürü olarak amaca uygun bir YBS algoritması seçilir.

3.2.1 Şekil Uzayı

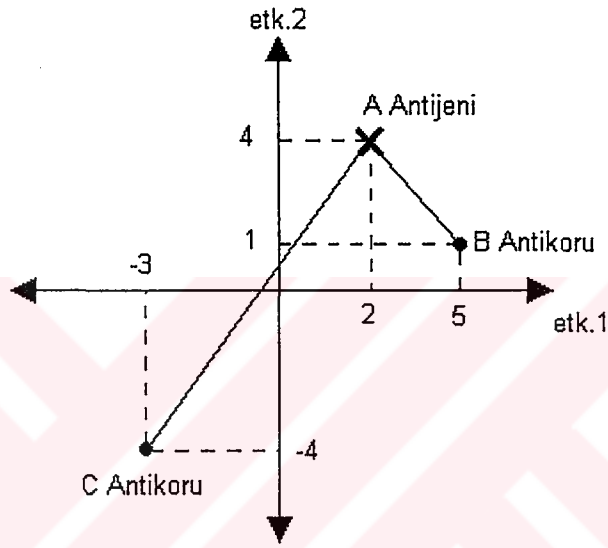
Bağışıklık elemanları arasındaki etkileşim pek çok fizikokimyasal olayı içerir. Mesela etkileşimdeki elemanlar arasında kovalent olmayan bağ, hidrojen bağı vb. bağların oluşması gibi. Bir antijenin bir antikör tarafından tanınması için, yüzeylerinin birbirlerini tamamlayacak şekilde bağlanmaları gerekir. Bu yüzden moleküller üzerinde geniş *tamamlayıcı bölgelere* ihtiyaç duyulur. Antijen ve antikörlerin birbirlerini tamamlamaları için gerekli olan şekil, elektrik dağılımları, uygun tamamlayıcı pozisyonları gibi özelliklerinin kümesine molekülün *genelleştirilmiş şekli* adı verilir.

Bir antikörün genelleştirilmiş şeklinin L tane parametre ile gösterilebildiğini varsayalım. Bu durumda L boyutlu bir uzayda bir nokta, antijenin bağlanma özelliklerine göre bağışıklık hücrelerinin üzerindeki moleküler reseptörlerin antijene bağlandıkların alanların *genelleştirilmiş şeklini* göstermiş olur ve *şekil uzayı-S* olarak adlandırılır.

Antijen tanıma için gerekli olan tüm hücrelerin oluşturduğu sete, *bağışıklık repertuarı* denir. Vücudumuzun tüm Antijenlerde korunabilmesi için bu repertuarın *tam* olması gereklidir. Vücut, herhangi bir Antijeni tanıyabilecek reseptör moleküllerini oluşturacak protein genlerine sahiptir ve bu genlerin farklı kombinasyonlarda dizilimi ile doğada bulunan hemen hemen tüm Antijenleri tanıyabilir. Yani vücudumuzun bağışıklık repertuarı *tamdır*.

Bunu bir örnek ile açıklayalım. Mesela bir canlının N boyutlu bir şekil uzayı olduğunu varsayalım. İlk başta da söylendiği gibi iki hücre birbirine bağlanırken

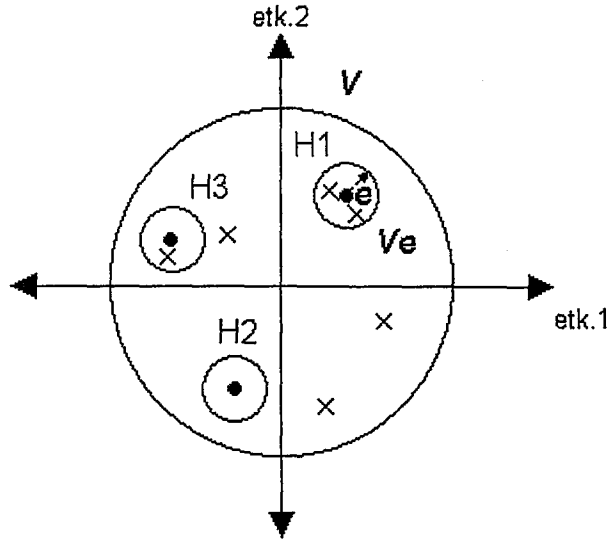
aralarında pek çok fizikokimyasal olay meydana gelir. İşte varsayılan bu N boyutlu şekil uzayında N tane eksen vardır ve her bir eksen, iki hücre arasında gerçekleşen fizikokimyasal etkileşimlerden birini temsil eder. Örneğin, bir A Antijeni ile bir B Antikoru arasında iki çeşit fizikokimyasal etkileşim gerçekleşebilir kabulü yapılsın. Bu durumda her bir fizikokimyasal etkileşimi temsilen 2 eksen kullanılacağı için iki boyutlu bir şekil uzayı, A Antijeni ile B ve C Antikorumu göstermek için kullanılabilir Şekil 3.9'da bu şekil uzayı gösterilmiştir.



Şekil 3.9: Şekil Uzayı Gösterimi

Şekil 3.9'da A Antijeni, B ve C Antikoru ve sahip oldukları fizikokimyasal özellikler gösterilmektedir. A antijeni ile C Antikoru arasındaki uzaklık, B Antikorundan fazla olduğu için, A Antijeni ile C Antikoru arasındaki etkileşimin şiddeti, A antijeni ile B Antikoru arasındaki etkileşimin şiddetinden daha fazladır diye gösterim yorumlanır.

Bağışıklık sisteminde antikorumların yabancı maddeleri tanıyabilmeleri için bu Antijenik yapılardan gelen uyarımların belirli bir eşik seviyesini aşması gerekmektedir. Bu özellik ise şekil uzayında tanıma çemberi adı verilen ve şekil 3.10'da görülen başka bir modelleme ile ifade edilmiştir.



Şekil 3.10 : Farklı bir şekil uzayı gösterimi (Tanıma çemberi)

Şekil 3.10'da üç tane antikor ve Antijenler iki boyutlu bir şekil uzayında gösterilmişlerdir. Her bir hücrenin tanıma çemberi e eşik seviyesi ile belirlenen V_e çemberidir. Şekilde H1 tanıma çemberi içinde kalan iki Antijeni, H3 ise sadece bir Antijeni tanır. H2'nin tanıma çemberinde herhangi bir Antijen olmadığı için H3 Antijen tanımamaktadır. Doğada mevcut olabilecek Antijen türleri, protein dizilimleri ile belirlendiği için belirli bir çeşitlilikte bulunabilirler. Şekil 3.10'deki iki boyutlu şekil uzayında bulunabilecek tüm Antijen türleri bir V hacmi ile ifade edilmiştir. Başka bir deyişle bulunabilecek tüm Antijenler mutlaka bu hacim içerisinde gösterilirler. Yine aynı şekil uzayında bulunan her bir Antikorum bir V_e tanıma hacmi olduğuna göre, şekil uzayında V hacmi içerisinde tüm noktaları kapsayacak biçimde bir Antikor populasyonu olduğunda bu Antikor populasyonu tüm Antijenleri tanıyabilir yani repertuar tamdır denebilir.

3.2.2 Duyarlılık Hesaplaması

Matematiksel olarak S şekil uzayındaki herhangi bir (m) molekülün genelleştirilmiş şekli L uzunluğunda bir nitelik stringi olarak gösterilebilir. Yani $m = \langle m_1, m_2, \dots, m_L \rangle$ stringi L boyutlu bir şekil uzayında bir noktadır. Bu string integer, gerçel, bitler veya sembollerden oluşabilir. Bu niteliklerin türü YBS

problemine göre seçilir. Bu tanımlamada önemli olan şey hangi ölçüt veya ölçütlerin niteliklerin etkileşim kalitesini yani duyarlılığını belirlemede kullanılacağıdır.

Duyarlılık genellikle bağışıklık birimlerinin birbirini tanıma kalitesini ifade eder. Çözülmesi beklenen problemin türüne göre duyarlılık hesabında kullanılacak yöntem belirlenebilir. Genellikle antijen ile antikor arasındaki uzaklığın hesaplanması kullanılır. Yukarıda da bahsedildiği gibi iki eleman birbirinden ne kadar uzak ise birbirlerini o kadar tamamlayacağı düşünülür. Elemanların gösterimine göre kullanılacak uzaklık ölçütü de değişir. Eğer elemanlar Öklid veya Manhattan uzaklığı kullanılır. Eğer antijenin ve antikorun ikili gösterimi söz konusu ise Hamming Uzaklığı kullanılır. Bu uzaklıkların formülize edilmiş halleri çizelge 3.1'de gösterildiği gibidir.

Antikor : $Ab = \langle Ab_1, Ab_2, \dots, Ab_L \rangle$;

Antijen : $Ag = \langle Ag_1, Ag_2, \dots, Ag_L \rangle$; şeklinde gösterildiği kabul edilsin.

Çizelge 3.1 : Duyarlılık Ölçütü için kullanılacak Uzaklıklar

Uzaklık Ölçütü	Formülü
Öklid Uzaklığı	$D = \sqrt{\sum_{i=1}^L (Ab_i - Ag_i)^2}$
Manhattan Uzaklığı	$D = \sqrt{\sum_{i=1}^L Ab_i - Ag_i }$
Hamming Uzaklığı	$D = \sum_{i=1}^L \delta \quad , \delta = \begin{cases} 1 & \text{eğer } Ab_i \neq Ag_i , \\ 0 & \text{diğer durumlarda.} \end{cases}$

3.2.3 Bağışıklık Algoritmaları

Bu bölümde YBS problemlerini çözmek için çeşitli bağışıklık mekanizmalarının modellenmesi için kullanılan algoritmalar tanıtılacaktır. YBS modellerini dört ana başlık altında toplayabiliriz.

- Kemik ilik Modeli (Bone marrow model) : bu model hücre ve moleküllerin repertuarlarının üretilmesi için kullanılırlar.
- Timmus Modeli (Thymus Model) :Self/nonself ayırımını yapabilme kabiliyetine sahip hücre ve moleküllerin repertuarlarının üretilmesi için kullanılırlar
- Klonal Seçme Algoritmaları (Clonal Selection Algorithms): Bağışıklık sistemi bileşenlerinin harici çevre ve antijenler ile olan etkileşimin nasıl olacağını kontrol eder.
- Bağışıklık Ağ Modelleri (Immune Network Models): Kendi yapılarını, dinamiklerini ve meta dinamiklerini içeren bağışıklık ağlarını modeller.

Bu bölümün geri kalanında tez için önemli olan son iki madde açıklanmıştır.

3.2.3.1 Klonal Seçme Algoritması

Klonal seçme adaptif bağışıklık sisteminin patojenik mikroorganizmalar ile nasıl ilgilendiğini anlatan bir teoridir. Bu teoriye göre prolife evresinde belli bir duyarlılıkla antijeni tanıyan B hücreleri seçilir. Bu hücreler aseksüel oldukları için çaprazlama olmaksızın mitoz bölünme ile çoğalırlar. Bu üreme evresi süresince B hücreleri hiper mutasyona tabi tutularak antijene karşı olan duyarlılıkları artırılır. Seçilen antijene olan duyarlılığın artırılması işlemi yani seçme ve mutasyon süreçlerine *bağışıklık cevabının duyarlılık olgunlaşması* adı verilir. B hücrelerindeki bu değişme işlemine ek olarak, antijenlere karşı yüksek duyarlılıkla aktive edilmiş B hücreleri, hafıza hücresi olmak üzere seçilirler. Uzun ömürlü olan bu hafıza hücreleri yüksek duyarlılık gösterdikleri antijenlerle veya benzerleriyle tekrar karşılaştıkları zaman daha çabuk cevap verilmesini sağlarlar.

Klonal seçme prensibinin iki temel özelliği bulunmaktadır:

- 1) *Duyarlılık Orantılı Üreme* : Antijene karşı yüksek duyarlılık gösteren seçilmiş b-hücrelerinden antijene olan duyarlılıkları ile orantı çoğaltılmasıdır.
- 2) *Mutasyon* : Her bir bağışıklık hücresinin üreme esnasında maruz kaldığı mutasyon oranı, duyarlılığı ile ters orantılıdır. Yüksek duyarlılığa sahip hücre az mutasyona

uğrarken duyarlılığı az olan hücre çok mutasyona uğrar. Bu olaya hiper mutasyon adı verilir.

Bu prensiplere dayanarak de Castro, L. N. ve Von Zuben tarafından 2000 yılında CLONALG isimli bir klonal seçme algoritması tanımlanmıştır. Öncelikli amaçları örüntü tanımak ve optimizasyon fonksiyonlarının çözülmesiydi.

CLONALG algoritmasının adımları aşağıdaki gibidir.

1. Başlangıç : Bireylerden rasgele bir başlangıç popülasyonu oluşturulur (P)
2. Antijen Sunumu : Her bir antijen (Ab) için, do
 - 2.1 Duyarlılık Hesaplanması : Ab antijeni P popülasyonuna sunulur ve popülasyondaki her bir birey ile olan duyarlılık değerleri belirlenir.
 - 2.2 Klonal Seçme ve Genişleme : P'den n_1 tane yüksek duyarlılıklı birey seçilir ve bu bireylerden antijene olan duyarlılıkları ile doğru orantılı olarak klon üretilir.
 - 2.3 Duyarlılık Olgunlaşması : Bu kopyaların tümü giriş antijeni ile gösterdikleri duyarlılığa ters orantılı olarak mutasyona uğratılır. Mutasyona uğramış bireyler P popülasyonuna eklenir ve yeniden en iyi bireyler seçilir. Bu bireyler M hafıza hücreleri olarak saklanır.
 - 2.4 MetaDinamik :Düşük duyarlılıklı n_2 tane birey rasgele üretilmiş yeni bireyler ile yer değiştirilir.
3. Döngü : Adım 2 durdurma kriteri ile karşılaşılıya kadar devam ettirilir.

3.2.3.2 Bağışıklık Ağ Modelleri

Bağışıklık sistemi için ilk önerilen yapıda sistem yabancı bir mikroorganizma tarafından uyarılmadığı sürece pasif durumdadır. Ancak antijenik bir yapı tarafında uyarıldığında dinamik hale geçer. Ancak daha sonradan sunulan bağışıklık ağ teorinse göre B hücre reseptörleri idiotop denilen bir yapıya sahiptir ve bu yapı sayesinde diğer serbest durumdaki antikörler veya reseptörler tarafından tanınabilir. Bağışıklık sisteminin elemanlarının tanıma ve tanınabilme yeteneği ile sistemin sürekli dinamik bir yapıda kaldığı söylenebilir. Bağışıklık ağ modelleri ise klonal

seçme prensibinden farklı olarak bağışıklık sisteminin her zaman dinamik olan bir yapısını modeller. Klonal seçme prensibi sadece antijenik bir yapı tarafından uyarıldığı zaman dinamik duruma geçtiği düşünülen sistemi modellerdi.

Bu tezde kullanılan bağışıklık ağ modeli aiNET algoritmasıdır. aiNET de Castro ve Von Zuben (2001) tarafından sunulmuş veri analizi, tanıma, sınıflandırma için kullanılabilen ağ tabanlı bir yapay bağışıklık algoritmasıdır. aiNET modeli antikor adı verilen ve birbirleri ile bağlantılı hücrelerden meydana gelir. Bu antikorların sisteme girebilecek antijenlerin dahili bir görüntüsünü oluşturduğu varsayılır. Bir sonraki kısımda aiNET algoritması daha ayrıntılı bir şekilde incelenecektir.

3.2.3.2.1 aiNET Algoritması

Bağışıklıksal modellemeyi tanımlayabilmek için tüm bağışıklıksal olayların bir S şekil uzayında gerçekleştiği farz edilir. Bu S şekil-uzayı, çok boyutlu ölçeklendirilmiş bir uzayıdır ve her eksen bir moleküler şeklin fiziko-kimyasal ölçütünü karakterize eder. Ayrıca şekil uzayının L boyutlu olduğu kabulü yapılmıştır. Bu şekil-uzayındaki her bir $s \in S$ noktası, Ab-Ab ya da Ag-Ab etkileşimlerini belirlemek için gerekli olan özelliklerin bir setini vermektedir (de Castro ve Timmis 2002).

Oluşturulacak ağda giriş örüntüleri Ag'lerdir. Ag-Ab ve Ab-Ab etkileşimleri benzerlik ölçütlerine dayalı olarak belirlenir. Bu benzerlik ölçütlerinin belirlenmesinde genellikle öklid uzaklığı kullanılır. Ag-Ab duyarlılığı, aralarındaki uzaklığın tersi ile orantılı olacaktır yani aralarındaki uzaklık az ise duyarlılık fazla, çok ise duyarlılık az olacaktır. Bu çalışmada ise kullanıcıların birbirleri ile olan benzerlikleri kullanılmıştır. Kullanıcılar arasında gerçek hayatta Ag-Ab arasında olan kilit anahtar ilişkisi olmadığı için öklid uzaklık ölçütü kullanılmamış onun yerine Pearson Bağıntısı ile benzerlikleri hesaplanmıştır. Yani benzerlik ile duyarlılık doğru orantılı hale gelmiştir.

Bu ağ yapısında iki farklı birim için duyarlılık hesaplanır. Bunlardan ilki, Ag_j - Ab_i çiftleri arasındaki duyarlılık ölçütü olan d_{ij} ; diğeri ise Ab_j - Ab_i çiftleri arasındaki duyarlılık ölçütü olan s_{ij} 'dir.

aiNET algoritmasını biraz daha detaylı inceleyelim. Bir t iterasyonunda aiNet'in N tane Ab içerdiği varsayılmıştır.

- Ab: N tane elementten oluşan mevcut Ab repertuarı ($Ab \ S^{NxL}$)
- $Ab_{\{m\}}$: toplam hafıza Ab repertuarı, m elementten oluşuyor ($Ab_{\{m\}} \ S^{mxL}$.
 $m \leq N$)
- Ag: M elemandan oluşan antijen popülasyonu ($Ag \ S^{MxL}$)
- f_j : tüm Ab 'lerin- Ab_i ($i=1, \dots, N$) Ag_j antijenine olan duyarlılıklarının bulunduğu vektör. Buradaki duyarlılıklar Ag-Ab uzaklığı ile ters orantılıdır.
- S: her Ab_i - Ab_j çifti arasındaki benzerlik matrisi . s_{ij} elementlerinden oluşuyor. ($i, j=1, \dots, N$)
- C: N_c tane elementten oluşan, Ab'lerden türetilen klon seti ($C \ S^{N_c x L}$)
- C^* : duyarlılık olgunlaşması (affinity maturation) işleminden sonraki C popülasyonu
- d_j : Ag_j antijeni ile C^* setinin her klonu arasındaki duyarlılıkları içeren vektör.
- ξ : seçilecek olgun Ab'lerin yüzdesi
- M_j : Ag_j antijeni için klonsal hafıza
- M_j^* : Ag_j antijeni için sonuçta oluşan klonsal hafıza
- σ_d : doğal ölüm eşiği
- σ_s : baskılama eşiği

aiNet öğrenme algoritmasının amacı, veriyi tanıyan ve onun yapısal organizasyonunu temsil eden bir hafıza seti oluşturmaktır. Bu kümenin doğruluğu ve ağın esnekliği ise antikor belirliliği ile kontrol edilir. Ab'ler ne kadar özel olurlarsa, ağdaki eleman sayısı da o kadar fazla olur. Antikor belirliliğini ise baskılama eşiği (σ_s) parametresi kontrol eder.

Yukarıda tanımlanan ifadelere dayalı olarak aiNet öğrenme algoritmasını daha detaylı incelersek:

(1.) Her iterasyon için, yap

(1.1) Her antijenik $Ag_j, j=1, \dots, M$ örüntüsü için:

(1.1.1) Bu antijenin tüm Ab_i 'lerle ($i=1, \dots, N$) olan $f_{i,j}$ duyarlılıklarını belirle:

$$f_{i,j} = 1/D_{i,j} \quad D_{i,j} = \|Ab_i - Ag_j\|, \quad i=1, \dots, N$$

(1.1.2) Bu duyarlılıklara bakarak, duyarlılıkları en yüksek olan n tane Ab 'yi seç ve bir $Ab_{(n)}$ alt seti oluştur.

(1.1.3) Seçilen bu n tane Ab , $f_{i,j}$ duyarlılıklarına bağlı olarak çoğalacaklar ve C klon seti oluşturacaklardır. Her Ab için, eğer duyarlılık daha yüksekse klon büyüklüğü de daha büyük olacaktır.

(1.1.4) Oluşturulan bu C klon setini, duyarlılık olgunlaştırma işlemine tabi tut ve olgunlaşmış bir C^* klon seti ortaya çıkar- C 'deki her k Ab 'si α_k oranı ile mutasyona uğrar. Bu α_k oranı da her Ab 'nin antijenik duyarlılığı $f_{i,j}$ ile ters orantılıdır. Yani yüksek duyarlıklı Ab 'ler düşük, az duyarlıklı Ab 'ler yüksek bir mutasyon oranı ile mutasyona uğrarlar:

$$C_k^* = C_k + \alpha_k (Ag_i - C_k); \quad \alpha_k \sim 1/f_{i,j};$$

$$k=1, \dots, N_c; i=1, \dots, N; j=1, \dots, M$$

(1.1.5) Ag_j antijeni ile C^* setinin her elemanı arasındaki duyarlılıkları $d_{k,j} = 1/D_{k,j}$ ile hesapla:

$$D_{k,j} = \|C_k^* - Ag_j\|, \quad k=1, \dots, N_c$$

(1.1.6) C^* seti içerisinde, en yüksek $d_{k,j}$ duyarlılığı olan Ab 'lerin % ξ 'sini seç ve M_j klonsal matrisine koy

(1.1.7) Apoptosis: M_j klonsal hafızasında $D_{k,j} > \sigma_d$ olan tüm elementleri ele. Yani Ag_j antijenine olan uzaklıkları doğal ölüm eşiği denilen σ_d 'den büyük olan tüm hafıza antikorlarını öldür.

(1.1.8) Ağ Etkileşimleri: Klonsal hafızanın elemanlarının her biri arasındaki benzerlik derecesi $s_{i,k}$ 'yi hesapla ($s_{i,k} = 0$ olduğunda benzerlik maksimumdur):
 $s_{i,k} = |M_{j,i} - M_{j,k}|$, $i,k=1, \dots, N_c$

(1.1.9) Klonsal baskılama: benzerlik dereceleri baskılama eşiğinden düşük olan (yani benzerlikleri fazla olan) - $s_{i,k} < \sigma_k$ tüm klonsal hafıza elemanlarını ele.

(1.1.10) Ağ oluşturma: Toplam Ab hafıza matrisi ile sonuçta oluşan M_j^* klonsal hafıza matrisini ardışık biçimde birleştir:
 $Ab_{\{m\}} = [Ab_{\{m\}}; M_j^*]$

(1.2) Ağ Etkileşimleri: $Ab_{\{m\}}$ içindeki tüm hafıza B'leri arasındaki benzerlik derecelerini belirle:

$$s_{i,k} = |Ab_{\{m\}}^i - Ab_{\{m\}}^k|, \quad i,k=1, \dots, m$$

(1.3) Ağ Baskılama: benzerlik dereceleri baskılama eşiğinden düşük olan ($s_{i,k} < \sigma_s$) tüm hafıza Ab'lerini ele

(1.4) Toplam Ab matrisini oluştur: $Ab = [Ab_{\{m\}}; Ab_{\{d\}}]$

(2) Durma şartını test et (de Castro ve Timmis 2002).

aiNET algoritmasının eğitimini durmak için aşağıdaki şartlardan birini sağlaması gerekmektedir.

1. Daha önceden belirlenmiş iterasyon sayısına ulaştığında
2. Daha önceden belirlenmiş antikor sayısına ulaştığında;
3. Antikorlar ile antijenler arasındaki ortalama hata istenen değerin altına düştüğünde
4. Birbirini takip eden k adım boyunca ortalama hata değeri belirlenen değerin altında kalıyorsa durdurma şartı sağlanmış olur. (de Castro ve Timmis 2002).

Ağ çıkışı olarak hafıza antikorlarının koordinatlarının yer aldığı $Ab_{\{m\}}$ matrisi ile , bu antikorlar arasındaki benzerlikleri barındıran bir S matrisi elde edilir. $Ab_{\{m\}}$ matrisi, aiNet'e sunulan Ag'lerin bir imajını temsil ederken, S matrisi ise hangi ağ

Ab'lerinin birbirleri ile bağlantılı olduklarını ve bu bağlantıların derecelerini temsil eder, başka bir deyişle genel ağ yapısını temsil etmiş olur.

Bundan sonraki adım, eğitilmiş aiNET'deki bilginin nasıl kullanılacağına karar verilmesidir. De Castro ve Von Zuben, eğitilmiş bir aiNet'i yorumlamak için kullanılabilir bir çok grafik konusu ve hiyerarşik kümeleme teknikleri sunmuşlardır. Bu çalışmada ise elde edilen hafıza matrisi S benzerlik matrisi kullanılarak K-Ortalama algoritması ile kümelendi.



4. YAPAY BAĞIŞIKLIK TABANLI BİR ÖNERİ SİSTEMİ MODELİ

Bu tezde yapay bağışıklık (Artificial Immune System) sistemi kullanılarak İFT tabanlı bir öneri sistemi tasarlanmıştır. Sistem tasarlanırken YBS kullanılmasının sebebi modellediği bağışıklık sisteminin az sayıda ve farklı gen kodlamalarına sahip bağışıklık elemanları ile çok sayıda ve farklı yapıdaki yabancı maddeleri tanıma özelliğine sahip olmasındandır. Ayrıca adaptif yapısı sayesinde bağışıklık sistemi vücuda daha önce hiç girmemiş ilk defa karşılaştığı mikropları da tanıyabilir. . Sunulan öneri modelinde veri tabanında bulunan kullanıcılar bağışıklık elemanlarını (antikorları) temsil eder ve oyladıkları filmler ile kodlanırlar. Sisteme yeni gelenler dışarıdaki yabancılar ise (antijenler) olarak kodlanır. Amaç mevcut antikorlardan, gelen antijenleri tanıyabilecek bir hafıza oluşturulmasıdır. Bunun için bir yapay bağışıklık algoritması olan AINET tabanlı bir model kurulmuştur. aiNET'in kullanılmasının sebebi öneri sistemlerinin başlıca sorunları olan bilgi eksikliği ve ölçeklenebilirlik problemlerinin her ikisinin de giderilebilecek potansiyele sahip olmasındandır. Bu bölümün devamında sırasıyla kullanılan "Movie Lens" veri kümesi tanıtılacak, kurulan aiNET tabanlı model açıklanacak ve uygulamadan elde edilen sonuçlar incelenecektir.

4.1 MovieLens Veri Kümesi ve Kullanıcıların Kodlanması

MovieLens, tavsiye sistemi araştırmaları için Minnesota Üniversitesi Bilgisayar Bilimleri ve Mühendisliği Bölümünde oluşturulan GroupLens isimli bir araştırma gurubu tarafından 1997'den beri www.movielens.org adresinden toplanan veriler ile oluşturulmuş araştırmacıların kullanımına açık bir veri kümesidir. Siteyi halen her hafta yüzlerce kullanıcı ziyaret etmekte, film oylayıp tavsiye almaktadır. Şu an sitenin 35000 kullanıcısı ve 3000'in üzerinde farklı film seçeneği bulunmaktadır.

MovieLens'in araştırmacıların kullanımına açık olan iki tane versiyonu bulunmaktadır. Bunlardan ilki 100,000 oy içeren *100K MovieLens* veri kümesidir. Bu oylar 1682 filmin 943 farklı kullanıcının oylaması sonucu elde edilmiştir. Diğer

kümesi ise *1Milyon MovieLens* veri kümesidir. Bunda ise 6040 kullanıcı tarafından oylanmış 3900 film için bir milyon oy bulunmaktadır. Her iki kümede de her bir kullanıcı en az yirmi ayrı film için oylama yapmıştır. Her bir kullanıcı oylayacağı filmi beğenisine göre bir ile beş arasında bir sayı ile değerlendirmiştir. Hiç oylanmamış bir filmin oy değeri ise sıfırdır. Ayrıca veri kümeleri kullanıcıya ait basit demografik (yaş,cinsiyet,yerleşim gibi) bilgileri de içerir.

Bu çalışma için 100K MovieLens kullanılmıştır. Veri kümesi %80'ni eğitim %20'si test olacak şekilde bölünmüştür. Kümedeki her bir kullanıcılarının kodlanması için SWAMI kodlaması kullanılmıştır (Cazyer ve Aickelin 2002). Bu kodlama aşağıda gösterildiği gibidir.

$$K = \{(FilmNo,Oy_1),(FilmNo,Oy_2),\dots,(FilmNo,Oy_N)\}$$

FilmNo : Filmin Numarası

Oy : Kullanıcının FilmNo ile temsil edilen filme verdiği oy

Bir kullanıcının komşulukları ise $Komsu = \{k_1,k_2,\dots,k_3\}$ şeklinde temsil edilir. Her bir $k_i \in K$ ise kullanıcıya komşu olan kullanıcıları gösterir.

4.2 Sunulan Öneri Modeli

aiNET de Castro ve Von Zuben (2001) tarafından sunulmuş veri analizi, tanıma, sınıflandırma için kullanılabilinen ağ tabanlı bir yapay bağışıklık algoritmasıdır. aiNET modeli antikor adı verilen ve birbirleri ile bağlantılı hücrelerden meydana gelir. Bu antikorların sisteme girebilecek antijenlerin dahili bir görüntüsünü oluşturduğu varsayılır. Antikorlardan inşa edilecek bir ağın aşağıdaki soruları cevaplaması beklenir. Verinin özü için herhangi bir grup ve ya alt grup var mı? Varsa kaç tane var? Bu verinin (grupların) yapısı veya görüntü dağılımı nasıldır? (de Castro ve Von Zuben 2001)

Yukarıdaki sorular ışığında bu çalışmada neden aiNET tabanlı bir model sunulduğunu açıklayalım. Öneri sistemleri için en büyük problemlerden birinin ölçeklenebilirlik olduğu ve bunun çözümü olarak da boyut azaltma tekniklerinin kullanılabilmesi yukarıda belirtilmişti. Bu sebepten verinin aslını elde etmeye yönelik bir algoritma olan aiNET çözüm için kullanılabilir. Bu sayede veri kümesi içindeki gruplar ortaya çıkacağından daha hızlı bir şekilde ve daha doğru komşulukların elde edilmesi sağlanır. Öneri sistemlerinin diğer bir sorunu olan bilgi eksikliğine de aiNET'in içinde barındırdığı hiper mutasyon mekanizması ile çözüm sunacağı düşünülmüştür. Mutasyon ile bilgi eksikliği her bir iterasyonda azaltıldığı ve tam bir veri kümesine doğru gidildiği için benzer kullanıcıların bulunması kolaylaşmış ve böylece bölüm 2.1.3.1'de anlatılan bilgi eksikliği sorununu gidermek için ayrı bir yöntem kullanma gereksinimi kalkmıştır. Görüldüğü üzere aiNET her iki problemin çözümünü de özünde barındırmakta ayrı ayrı işlem yapma zahmetini ortadan kaldırmaktadır.

Aşağıda bölüm 3.2.3.2.1' de ayrıntılı bir şekilde anlatılan aiNET algoritmasının öneri sisteminde nasıl kullanıldığı açıklanacaktır. Bağışıklık sistemi ile Öneri sisteminin modellenen birimleri arasındaki etkileşim aşağıda görüldüğü gibidir.

- **Ab** : Eğitim kümesinde bulunan kodlanmış kullanıcılarıdır.
- **Ab_{m}** : Hafıza Matrisine alınmış kullanıcılar
- **Ag** : Test kümesinde bulunan kodlanmış kullanıcılarıdır.
- **f_j** : **Ab_i** (i=1,...N) ile **Ag_j** arasındaki duyarlılığın tutulduğu vektör. Benzerlik için Pearson bağıntısı kullanılır. Benzerlik ile duyarlılık doğru orantılı alınır.
- **Mutasyon Yöntemi:** Saf Bayes Sınıflandırıcısı kullanılmıştır.

Tasarlanan modelde tavsiye almak için siteye gelen kullanıcılar antijenler olarak görülür. Yeni gelen kullanıcıdan öncelikle kendine sunulan filmler arasından en az 20 tanesini oylaması istenir. Bu oylama bilgileri kullanılarak kullanıcı 3.1'de anlatıldığı şekilde kodlanır ve **Ag_i** olarak sisteme alınır. Daha önceden veritabanında

oylama bilgileri bulunan kullanıcılar ise Antikor kümesini oluşturur. Benzerlik ölçütü olarak formül 2.1 ile verilen pearson bağıntısı kullanılmıştır. Mutasyon yöntemi olarak ise şimdiye kadar yapılan çalışmalarda kullanılan tek noktaya göre mutasyon, çoklu noktaya göre mutasyon vb. yöntemlerden farklı olarak o anki bireye bağlı rasgele bir değer değişimi yerine geçmişteki bireylerin değerlerine de bağlı olarak olasılıksal bir değer üreten Saf Bayes Sınıflandırıcı kullanılmıştır. Bireyin daha önce oylamadığı bir film rasgele seçilmiş, ona saf bayes sınıflandırıcısı kullanılarak o filme verebileceği oy hesaplanmış ve kullanıcıyı temsil eden vektöre eklenmiştir. Böylece sistem kendi içerisinde eğitimine devam ederken bilgi eksikliğini azaltmış, her adımda daha doğru benzerliklerin hesaplanmasını sağlamıştır.

Aşağıda aiNET öğrenme algoritmasının sunulan öneri sistemi için adımları verilmiştir. Görüldüğü üzere bu algoritmada iki tür baskılama söz konusudur. Bunlar sırasıyla Ağ baskılama ve Klonal baskılamadır.(1.1.9 ve 1.3) Öneri sistemi için aiNET’i cazip kılan noktalardan birisi bu baskılama mekanizmasıdır. Klonal baskılama ile her bir antijene yani aktif kullanıcıya uygun olmayan antikorlar elenir. Her bir antijen için seçilen antikorlar ise hafıza matrisinde birleştirilir. Ancak bu matriste birbirleri ile benzer antikorlar bulunabilir. Birinin tanıdığı antijeni diğeri zaten tanıyacağı için buna gerek yoktur. Bu aşamada Ağ baskılaması devreye girer ve birbirleri ile olan benzerlikleri belli bir değerin üzerinde olan antikorlar hafızadan elenir. Böylelikle ölçeklenebilirlik problemi aşılmış, sonuçta elde edilen matris ise test kullanıcıları için dahili bir hafıza niteliği kazanmıştır. Tavsiye ve tahmin üretmede bu hafızanın kullanılması komşulukların bulunmasını kolaylaştırdığından ve doğru komşu seçimini artırdığından sistemin performansına olumlu etkisi olmuştur.

aiNET öğrenme algoritmasının adımları sunulan model için aşağıdaki gibidir.

1. Her bir iterasyon için, do

1.1 Test kümesindeki her bir antijenik örüntü için $Ag_j, j=1, \dots, M$ ($Ag_j \in Ag$), do

- 1.1.1 Ag_j 'nin her bir Ab_i ile olan benzerliğini bul. f_{ij} vektörünü oluştur. Duyarlılık hesaplamasında formül 2.1'deki Pearson bağıntısını kullan.
 $f_{ij} = \text{Benzerlik}(Ag_j, Ab_i), i=1, \dots, N$
- 1.1.2 En yüksek duyarlılık değerine sahip n tane antikoru seç. $Ab_{\{n\}}$ alt kümesini oluştur. (Komşulukları belirle)
- 1.1.3 Seçilen bu n antikoru duyarlılıkları ile doğru orantılı olarak klonla ve C kümesi oluştur. (Yüksek duyarlılığa sahip antikoru daha çok klonlanırken düşük antikora sahip olan daha az klonlanır).
- 1.1.4 C kümesini duyarlılık olgunlaşması sürecine gönder. Mutasyona uğradıktan sonra C^* kümesi oluştur. (Her bir antikoru duyarlılığı ile ters orantılı olarak mutasyona uğrar. Mutasyon için saf Bayes sınıflandırıcısı kullanılmıştır.)
- 1.1.5 C^* kümesi elemanları ile Ag_j arasındaki duyarlılıkları hesapla ve d_{kj} vektörü oluştur.
 $d_{kj} = \text{Benzerlik}(C^*_k, Ag_j), k = 1, \dots, N_c$
- 1.1.6 C^* kümesinden en yüksek d_{kj} değerine sahip antikorklardan $\% \zeta$ kadarı tekrar seç ve klonal hafızanın M_j matrisine koy.
- 1.1.7 Ölüm (Apoptosis): Duyarlılığı belli bir değerin altında olan hafıza elemanları M_j matrisinden çıkar.
- 1.1.8 Hafıza klonları arasındaki duyarlılık $s_{i,k}$ 'yi hesapla. $s_{i,k} = \text{Benzerlik}(M_{j,i}, M_{j,k}), \forall i, k$
- 1.1.9 Klonal Baskılama : σ_s 'den büyük olan $s_{i,k}$ değerine sahip hafıza hücrelerini ele. (M_j^* matrisi elde edilir.)

1.1.10 Ag_j için bileşik klonal hafıza M_j * matrisini,
toplam Hafıza antikor matrisine $Ab_{\{m\}}$ 'ye ekle.

$$Ab_{\{m\}} \leftarrow [Ab_{\{m\}} ; M_j *]$$

1.2 $Ab_{\{m\}}$ içindeki tüm hafıza antikorları arasındaki benzerlik derecelerini belirle.

$$s_{i,k} = \text{Benzerlik}(Ab_{\{m\}}^i, Ab_{\{m\}}^k) \quad i, k = 1, \dots, m$$

1.3 Ağ baskılama: Benzerlik dereceleri baskılama eşiğinden yüksek olan ($s_{i,k} > \sigma_s$) tüm hafıza Ab'lerini ele.

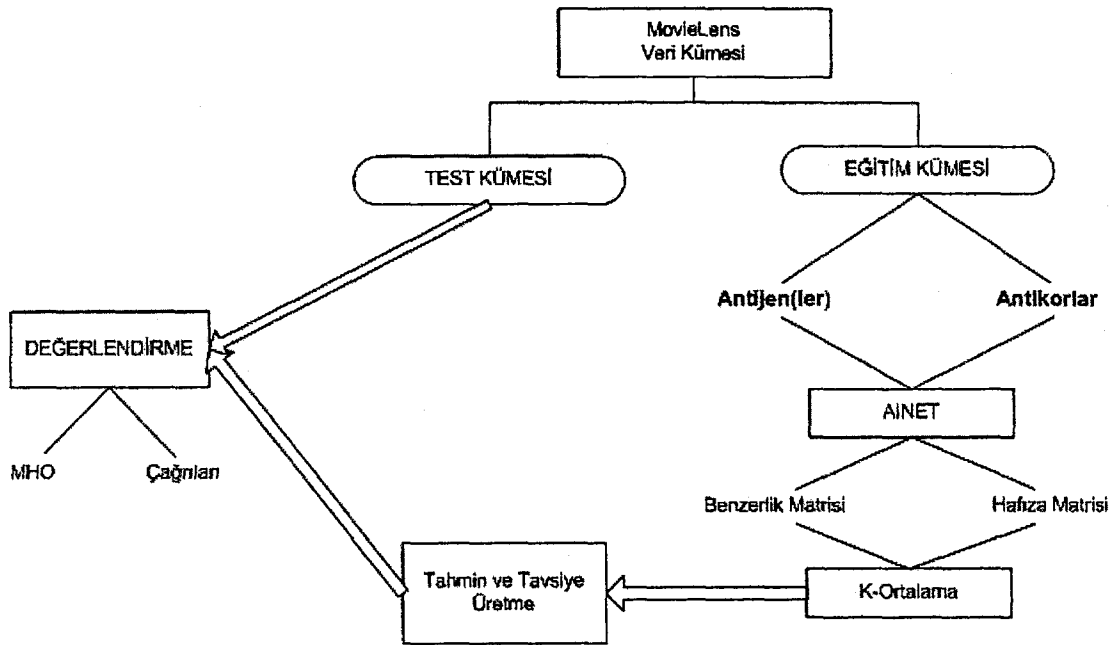
2 Durdurma şartını test et.

4.3 Uygulama

Bu bölümde tez için yazılmış uygulamanın çalışma mantığı, uygun parametre değerlerinin bulunması için yapılan analizler ve elde edilen sonuçlar ele alınmıştır.

4.3.1 Çalışma Mantığı

Bölüm 4.1'de açıklanan 100,000 oylu veri kümesi %80'ni eğitim, %20 si test olmak üzere ikiye ayrılmıştır. Eğitim kümesindeki bir yada birden fazla kullanıcı antijen olarak seçilmiş ve Ag matrisine, geriye kalan kullanıcılar ise Ab matrisine atanmıştır. Giriş matrisleri belli olduktan ve giriş parametrelerine de değerleri atandıktan sonra eğitime başlanmıştır. Algoritmanın çalışması sonucunda da hafıza antikorlarının ve benzerliklerin bulunduğu iki matris elde edilmiştir. Daha öncede belirtildiği gibi hafıza matrisi direk bilgi çıkarımında kullanılamıyordu. Bu yüzden elde edilen hafıza matrisi K-Ortalama algoritması kullanılarak kümelenecek şekilde komşulukların oluşması da sağlanmıştır. Daha sonra bölüm 2.1.2.2'de bahsedilen yöntemler kullanılarak tahmin ve tavsiye üretilmiştir. Hesaplan bu tahmin ve tavsiyelerin değerlendirilmesi için test kümesinden antijen olarak seçilen kullanıcılara karşılık gelen oylanmış filmlerin listesi alınmış ve 2.1.2.3'de anlatılan MHO ve Çağrılan değerlendirme ölçütleri kullanılarak hata hesabı yapılmıştır. Sunulan modelin yapısı şekil 4.1'de gösterildiği gibidir.



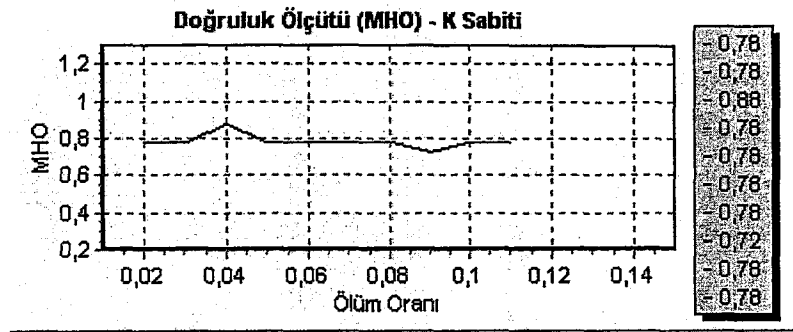
Şekil 4.1 : aiNet Tabanlı Öneri Sistemi

4.3.2 Uygun Parametre Değerlerinin Tespiti

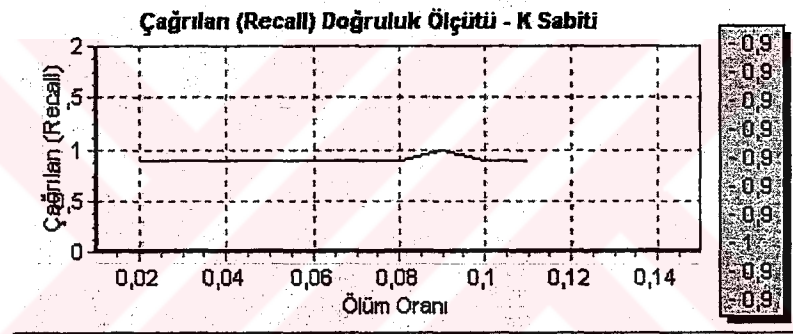
aiNet tabanlı öneri sistemini etkileyen kullanıcı bağımlı 4 tane parametre bulunmaktadır. Bunlar doğal ölüm eşiği (apoptosis threshold), baskılama eşiği (suppression threshold), ζ yüzdesi ve n sayısıdır. Bu parametreler için en uygun değerlerin seçilmesi modelin performansını etkileyen en önemli faktörlerdendir. Bu yüzden her bir parametre için, diğerleri sabit tutulurken seçilen parametrenin değeri belirtilen aralıkta değiştirilerek model çalıştırılmıştır. Sırası ile incelersek;

- *Doğal ölüm Eşiği (Apoptosis threshold)*: Duyarlılığı bu eşik değerinin altında olan hafıza elemanları M_j hafıza matrisinden çıkarılır. Böylece antijen ile benzerliği istenenin altında olan bireylerin işleme katılması engellenip işlem süresi kısaltılır. Değeri 0,01 ile 0.11 aralığında değiştirilerek model çalıştırılmıştır. Bu esnada diğer parametrelerin değeri *baskılama_eşiği*=0.01 ; *ζ yüzdesi*=%50; *n*=20 olarak set edilmiştir. Elde edilen ilk grafik *ölüm_eşiği* değerine göre MHO değerlendirme ölçütünün değerinin değişimini

göstermektedir. MHO değerinin en düşük olduğu değer sistemin en az hata ile tahmin ürettiği değerdir. Grafikten bu değer 0.09 olduğu görülmektedir.



Şekil 4.2 : ölüm_eşiği değerine göre MHO değerinin değişimi

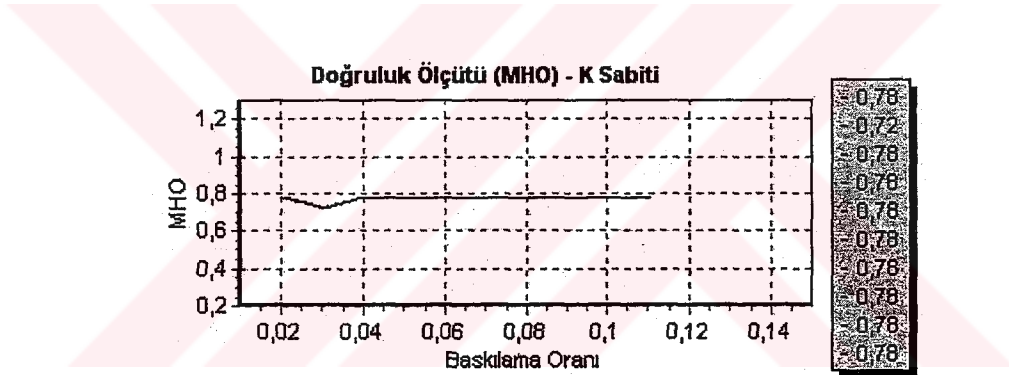


Şekil 4.3 : ölüm_eşiği değerine göre Çağrılan değerinin değişimi

Değerlendirme de kullanılan diğer bir ölçüt ise çağrılan parametresidir. Yine ölüm_eşiği parametresinin değeri $0,01$ ile $0,11$ aralığında değiştirilerek model çalıştırılmıştır. Bu esnada diğer parametrelerin değeri *baskılama_eşiği*= 0.01 ; ζ yüzdesi=%50; $n=20$ olarak set edilmiştir. Şekil 4.3'de gösterilen grafik ölüm_eşiği değerine göre Çağrılan değerlendirme ölçütünün değerinin değişimini göstermektedir. Çağrılan değeri ne kadar yüksekse verilen tavsiyenin doğruluğu o kadar fazla olacaktır. Şekil 4.2'de ki grafikten MHO için parametrenin en uygun değerinin $0,09$ olduğu görülmüştü. Şekil 4.3'deki grafikten bu değere karşılık gelen Çağrılan değerine bakılırsa 1 olduğu görülür. Bunun anlamı sistem test kümesindeki sistemden saklanan tüm

filmleri tavsiye listesinde göstermiş demektir. Yani %100 bir başarı elde edilmiştir.

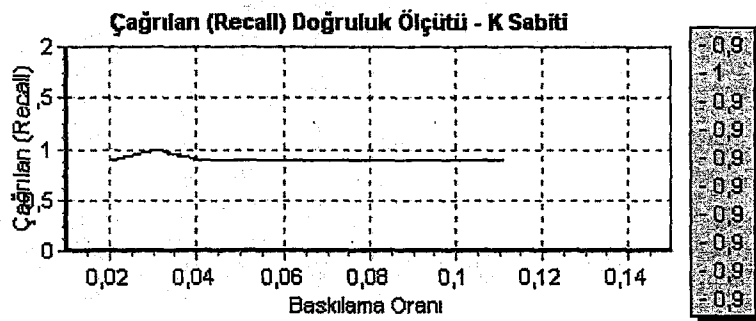
- *Ağ Baskılama Eşiği (Network Suppression threshold)*: Benzerlik dereceleri baskılama eşiğinden yüksek olan ($s_{i,k} > \sigma_s$) tüm hafıza Ab'leri elenir. Böylece ağda benzer antikorların çoğalması engellenerek hafıza matrisinin gereksiz şekilde büyümesinin önüne geçilmiş olunur. Değeri 0,01 ile 0,11 aralığında değiştirilerek model çalıştırılmıştır. Bu esnada diğer parametrelerin değeri $\text{ölüm_eşiği}=0.01$; ζ yüzdesi=%50; $n=20$ olarak set edilmiştir. Elde edilen ilk grafik *ağ_baskılama_eşiği* değerine göre MHO değerlendirme ölçütünün değerinin değişimini göstermektedir. MHO değerinin en düşük olduğu değer sistemin en az hata ile tahmin ürettiği değerdir. Grafikten bu değer 0.03 olduğu görülmektedir.



Şekil 4.4 :*ağ_baskılama_eşiği* değerine göre MHO değerinin değişimi

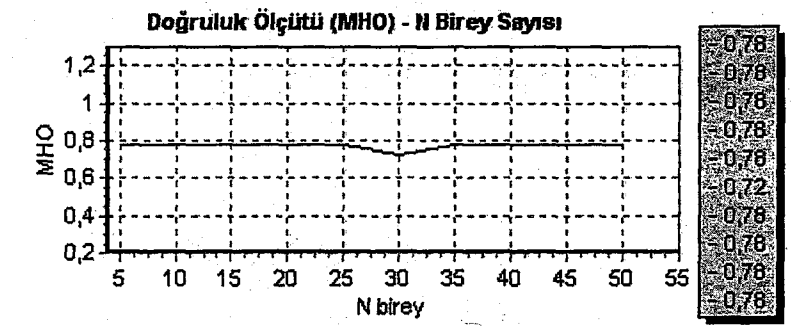
ağ_baskılama_eşiği parametresinin değeri 0,01 ile 0,11 aralığında değiştirilerek model tekrar çalıştırılmış ve çağrılan değerlendirme ölçütüne göre değerlendirilmiştir. Bu esnada diğer parametrelerin değeri $\text{ölüm_eşiği}=0.01$; ζ yüzdesi=%50; $n=20$ olarak set edilmiştir. Şekil 4.5'de gösterilen grafik *ağ_baskılama_eşiği* değerine göre Çağrılan değerlendirme ölçütünün değerinin değişimini göstermektedir. Şekil 4.4'de ki grafikten MHO için parametrenin en uygun değerinin $0,03$ olduğu görülmüştü. Şekil 4.3'deki grafikten bu değere karşılık gelen Çağırın değerine bakılırsa 1 olduğu görülür. Aynı bir önceki analizde olduğu gibi sistem uygun değerde

test kümesindeki sistemden saklanan tüm filmleri tavsiye listesinde göstermiştir. Yani %100 bir başarı elde edilmiştir.



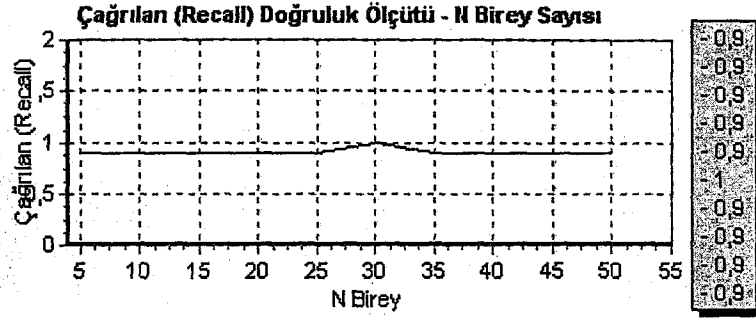
Şekil 4.5 : $ağ_baskılama_eşiği$ değerine göre Çağrılan değerinin değişimi

- N birey sayısı : Ag_i 'nin ilk adımda Ab matrisi ile olan benzerliği hesaplandıktan sonra oluşan f_{ij} vektöründen en yüksek değerlikli n birey seçilir ve klonlanır. Bu parametre seçilen n birey sayısını belirler. Bu parametrenin analizinde de bir önceki analizlerde olduğu gibi diğer parametrelerin değerleri sabitlenir ve n değeri belli bir aralıkta değiştirilir. $ölüm_eşiği=0.01$; $baskılama_eşiği=0.01$; $\zeta=\%50$ değerleri atanırken n değeri 5 ile 55 arasında değiştirilmiş ve aşağıdaki grafikler elde edilmiştir.



Şekil 4.6 : n birey değerine göre MHO değerinin değişimi

Şekil 4.6'de n değeri ile MHO değerinin değişimi gösterilmiştir. Bu grafikten bu parametre için en uygun değer 30 olduğu görülmektedir.

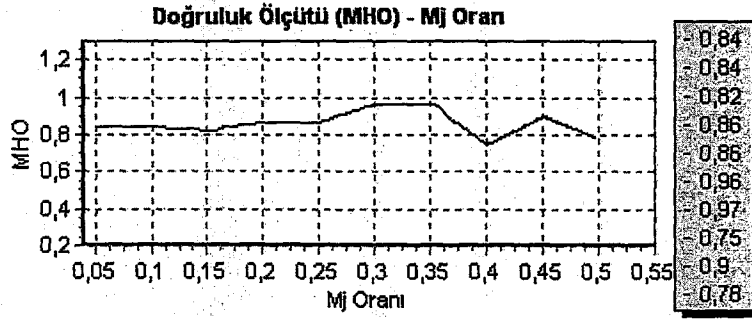


Şekil 4.7 : n birey değerine göre Çağrılan değerinin değişimi

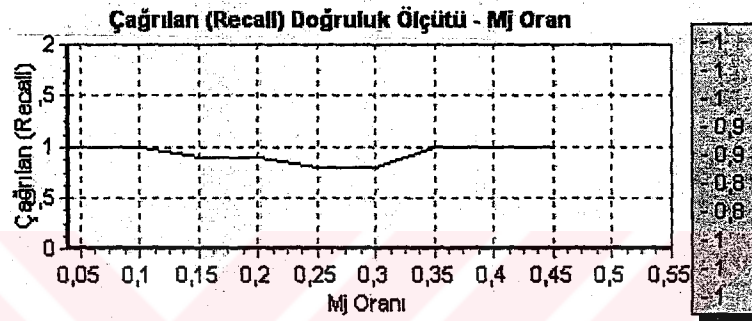
Şekil 4.7’de ise n değeri ile çağrılan değerinin değişimi gösterilmiştir. Burada n ‘in 30 birey değeri için çağrılanın 1’e eşit olduğu görülmektedir. Bu değer bir önceki grafikte de uyumludur yani en düşük MHO değerine karşılık gelmektedir.

- ζ yüzdesi : Klonlanlandıktan sonra oluşan C matrisinin elemanları ile Antijen arasında tekrar duyarlılık hesaplanır. Bu hesabın sonunda duyarlılıklar en yüksek olan elemanları $\% \zeta$ hafıza matrisine atanır. Hafıza matrisinin boyutunu dolayısı ile veri yapısının dahili görüntünün boyutunu belirlediği için önemli bir parametredir. Bu parametrenin analizinde de bir önceki analizlerde olduğu gibi diğer parametrelerin değerleri sabitlenir ζ değeri belli bir aralıkta değiştirilir. $\text{ölüm_eşiği}=0.01$; $\text{baskılama_eşiği}=0.01$; $n=20$ değerleri atanırken ζ değeri 0,05 ile 0,5 arasında değiştirilmiş ve aşağıdaki grafikler elde edilmiştir.

Şekil 4.8’da ζ değeri ile MHO değerinin değişimi gösterilmiştir. Bu grafikten bu parametre için en uygun değer 0.4 olduğu görülmektedir.



Şekil 4.8 : ζ değerine göre MHO değerinin değişimi



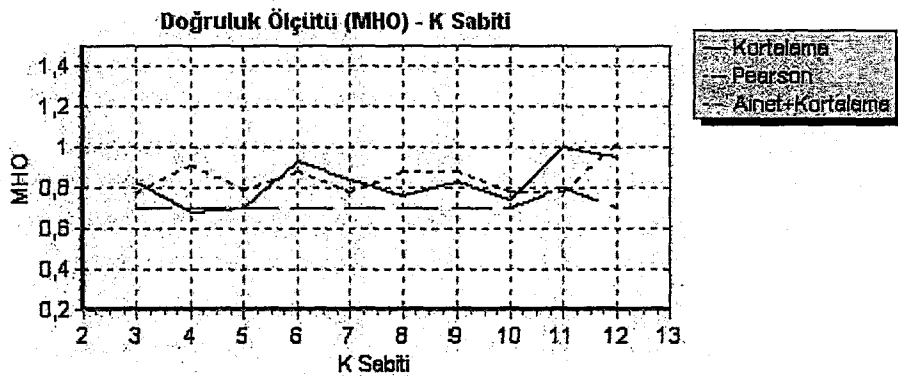
Şekil 4.9 : ζ değerine göre Çağrılan değerinin değişimi

Şekil 4.9'de ise ζ değeri ile çağrılan değerinin değişimi gösterilmiştir. Burada ζ 'ın 0,05;0,1;0,35;0,4;0,45 değerlerinde çağrılanın 1'e eşit olduğu görülmektedir. Ancak bir önceki analizde en iyi sonucu 0,4 verdiği için ζ parametresi için en uygun değer olarak %40 seçilir.

4.3.3 Sunulan Modelin Diğer Yöntemlerle Kıyaslanması

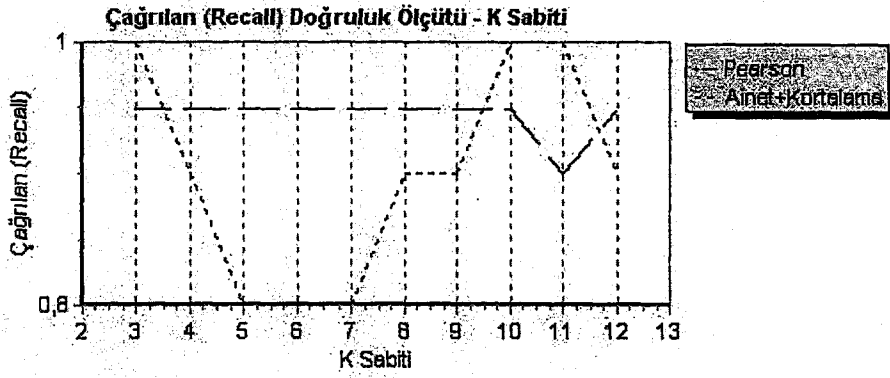
Sunulan model için parametrelerin en uygun değerleri bir önceki bölümde yapılan analizler sonucunda belirlenmiştir. Yani $\text{ölüm_eşiği}=0.09$; $n=30$; $\zeta=\%40$ ve $\text{baskılama_eşiği}=0.03$ olarak seçilmiş ve ağ eğitilmiştir. Elde edilen sonuçlar literatürde öneri sistemleri için kullanılan k-ortalama ve standart İFT yöntemleri ile kıyaslanmıştır. Hesaplanan tahmin ve tavsiye değerlerini değerlendirmek için MHO ve çağrılan ölçütleri kullanılmıştır ve bunların k sabitine göre değişimleri Şekil 4.10

ve şekil 4.11 de gösterilmiştir. Şekil 4.12’de ise aiNet ve Standart İFT yönteminin tavsiye zamanınının k sabitine göre değişimi gösterilmiştir.



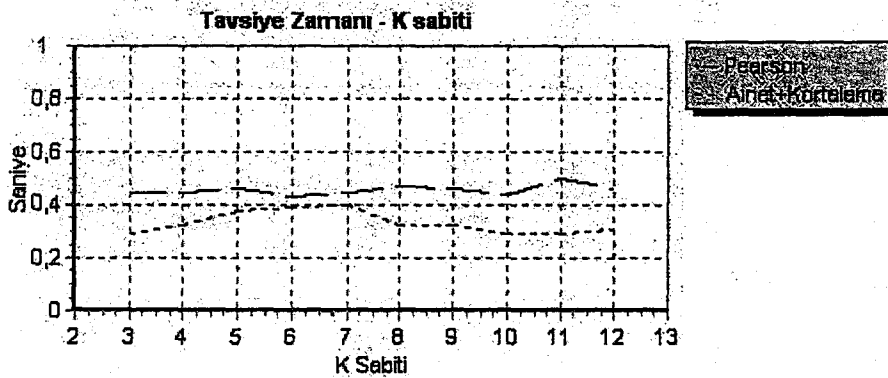
Şekil 4.10 aiNet, k-ortalama, standart İFT yöntemlerinin MHO ölçüt değerlerinin k sabitine göre değişimi

Öneri sistemi için üç yöntem yazılmış ve denenmiştir. Bunlardan ilki sunulan yöntem olan aiNET algoritmasıdır. aiNET sonucu elde edilen hafıza matrisinin kullanılabilmesi için matrise k-ortalama uygulanmış ve kümelenebilirliği sağlanmıştır. Diğer yöntemler ise veriye direkt k-ortalama ve standart İFT algoritmalarının uygulanmasıdır. İFT algoritması için kullanılacak komşuluk sayıları, aiNET ile kıyaslayacağımız için, hafıza kümesine k-ortalama algoritması uygulandıktan sonra her bir kümede olan kullanıcı sayısına karşılık gelmektedir. Komşulukların oluşturulmasında bölüm 2.1.2.1’de anlatılan *kümeleme yardımı ile komşuluk seçimi* yöntemi kullanılmıştır. Hesaplanan tahmin ve tavsiye değerlerinin değerlendirilmesi için MHO ölçütü kullanılmış ve şekil 4.10’daki grafik elde edilmiştir. Bu grafiğin oluşturulmasındaki amaç hem yöntemlerin kıyaslanması hem de model için en uygun k sabitinin belirlenmesidir. Grafikten görüldüğü üzere en uygun k değeri 10 ve ya 11 ‘dir. Ayrıca bu değerler için algoritma beklenen hata aralığında değer üretmiştir.



Şekil 4.11 aiNet ve Standart İFT yönteminin çağrılan ölçüt değerlerinin k sabitine göre değişimi

Şekil 4.11' de aiNet ve Standart İFT yönteminin çağrılan ölçüt değerlerinin k sabitine göre değişimi gösterilmektedir. Grafikte kesikli çizgi standart İFT algoritmasını noktalı çizgi ise aiNET'i göstermektedir. Bu grafikte, şekil 4.10'daki grafikten elde edilen aiNET için en uygun k değerleri olan 10 ile 11 aralığını göz önüne alınır. Bu aralıkta sunulan modelin %100 başarı ile tavsiye ürettiği görülmektedir. Buna karşın standart İFT algoritması %90 ila %95 doğrulukla tavsiye üretebilmiştir. aiNET tabanlı yöntem daha iyi sonuç vermiştir.



Şekil 4.12 aiNet ve Standart İFT yönteminin tavsiye zamanının k sabitine göre değişimi

Diđer bir deęerlendirme yntemi de tavsiye verme hızıdır. Őekil 4.12'de aiNet tabanlı model ile Standart İFT ynteminin tavsiye zamanının k sabitine gre deęişimi gsterilmiştir. Grafikten de grldę zere sunulan yntem daha hızlı tavsiye vermektedir.



5.SONUÇLAR

Yapay bağımsızlık sisteminde yaşanan gelişmeler iki farklı alanda ama birbirleri ile ilişkili bir şekilde devam etmektedir. Bunlardan ilki yeni algoritmalar ve modeller sunularak bu araştırma sahasının geliştirilmesini sağlamaktır. Diğer ise mevcut modelleri ve algoritmaları gerçek hayattaki sorunlara uygulayıp kullanılabilirliği artırmaktır. Bu çalışmada ikinci alanda yer almaktadır. Amacı daha verimli bir öneri sisteminin oluşturulmasıdır. Bunun içinde aiNET tabanlı bir model sunulmuştur.

aiNET tabanlı model için uygun parametre değerleri seçildiğinde modelin diğer yöntemler ile benzer ve beklenen hata seviyesinde tahmin ürettiği gözlemlenmiştir. Bölüm 3.3.3 'de gösterildiği gibi standart İFT en iyi %80 doğrulukta tavsiye verebilirken önerilen model uygun k sabiti alındığında %100 doğrulukta tavsiye verdiği görülmüştür. Ayrıca tavsiye verme hızında da iyileştirme sağlanmıştır. Sonuçta öneri sistemleri için müşteri sadakatinin ve güveninin temini için hayati önemi olan verilen önerinin doğruluğunun artması ve öneri verme süresinin ise kısılması sağlanmış ve diğer yöntemlere göre daha avantajlı olduğu yapılan analizler sonucunda gösterilmiştir.

6. KAYNAKLAR

- Breese J.S., Heckerman D., Kadie C.,** 1998. Empirical analysis of predictive algorithms for collaborative filtering, Proceedings of 14th Conference on Uncertainty in Artificial Intelligence, 43-52.
- Blum A. ,**1997. Empirical Support for Winnow and Weighted-Majority Algorithms: Results on a Calendar Scheduling Domain, Machine Learning, 26,1,5-23.
- Cazyer S., Aickelin U.** 2002. Recommender system Based on the Immune Network, Proceeding CEC2002, 807-813.
- Chen Q., Aickelin U.** 2004. Movie Recommendation System using an Artificial Immune System, Poster Proceedings of ACDM.
- Cho Y.H., Kim J.K.** 2004. Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce, Expert Systems with Applications, 26, 233-246.
- de Castro, L. N. , Von Zuben, F. J.** 2001. aiNet: An Artificial Immune Network for Data Analysis, Data Mining: A Heuristic Approach kitabındaki XII bölüm, H. A. Abbass, R. A. Sarker, and C. S. Newton (eds.), Idea Group Publishing, USA, 231-259.
- de Castro, L. N., Timmis, J. I.** 2002. Artificial Immune Systems: A New Computational Intelligence Approach, Springer -Verlag, London, 357 sayfa.
- de Castro, L. N., Von Zuben, F. J. ,** 1999. Artificial Immune Systems: Part I – Basic Theory and Applications, Teknik Rapor – RT DCA 01/99, 95 sayfa.
- de Castro, L. N., Von Zuben, F. J.** 2000. Artificial Immune Systems: Part II – A Survey of Applications, Teknik Rapor – RT DCA 02/00, 65 sayfa.

- de Castro, L. N., Von Zuben, F. J.** 2002. Learning and Optimization Using the Clonal Selection Principle, IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems, 6,3. 239-251.
- Han J., Kamber M.** 2001. Data Mining: Concepts and Techniques , Morgan Kaufman Publishers , 550 sayfa.
- Herlocker, J.,** 2000. Understanding and Improving Automated Collaborative Filtering Systems., Ph.D Dissertation, University of Minnesota.
- Li Q., Kim B.M.,** 2003. Clustering approach for hybrid recommender system, Proceedings. IEEE/WIC International Conference - Web Intelligence.
- Miller B. N.,** 2003. Toward a personal recommender system, PhD thesis, University of Minnesota.
- Pazzini M. J.,** 1999. A Framework for Collaborative, Content Based and Demographic Filtering, Artificial Intelligence Review, 13(5-6) 393-408.
- Resnick P., Lacovou N., Suchak M., Bergstrom P., Riedl J.,** 1994. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In proceedings of CSCW94.
- Sarwar B.M., Karypis G. Kotsan J.A, Riedl J.T.** 2002. Recommender Systems for Large-scale E-Commerce: Scalable Neighborhood Formation Using Clustering, Proceedings of the Fifth International Conference on Computer and Information Technology (ICIT 2002).
- Sarwar B.M.** 2001. Sparsity, Scalability and Distribution in Recommender Systems, PhD Thesis, University of Minnesota.
- Sarwar B.M., Karypis G. Kotsan J.A, Riedl J.T.** 2001. Item-based Collaborative Filtering Recommendation Algorithms, WWW10, Hong Kong.

Sarwar B.M., Karypis G. Kotsan J.A, Riedl J.T. 2000. Application of
Dimensionality Reduction in Recommender System – A Case Study.

