İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

SOSYAL AĞ VERİSİ KULLANILARAK MOBİL TELEFONLAR İÇİN ÖNERİ ALTYAPISI TASARLANMASI

YÜKSEK LİSANS TEZİ
Alper ÖZCAN

Anabilim Dalı : Bilgisayar Mühendisliği

Programı : Bilgisayar Mühendisliği

Tez Danışmanı: Yrd. Doç. Dr. Şule Gündüz ÖĞÜDÜCÜ

HAZİRAN 2010

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

SOSYAL AĞ VERİSİ KULLANILARAK MOBİL TELEFONLAR İÇİN ÖNERİ ALTYAPISI TASARLANMASI

YÜKSEK LİSANS TEZİ
Alper ÖZCAN
(504071505)

Tezin Enstitüye Verildiği Tarih : 07 Mayıs 2010
Tezin Savunulduğu Tarih : 14 Haziran 2010

Tez Danışmanı : Yrd. Doç. Dr. Şule Gündüz ÖĞÜDÜCÜ (İTÜ)
Diğer Jüri Üyeleri : Doç. Dr. Zehra ÇATALTEPE (İTÜ)
Yrd. Doç. Dr. Banu DİRİ (YTÜ)

HAZİRAN 2010

Anneme ve babama,

ÖNSÖZ

Gerek yüksek lisans ders aşamasında; gerek tez çalışması sırasında desteğini, önerilerini ve yol göstermeyi esirgemeyen danışmanım Yrd. Doç. Dr. Şule Gündüz ÖĞÜDÜCÜ'ye öncelikli olarak teşekkür ederim. Benim için yeni olan birçok konuda hem sorularıyla hem de ilettikleriyle bana ışık tutmuştur.

Hayatımın her aşamasında bana güvenen ve desteğini hiç esirgemeyen aileme ayrıca hem sevgisini, hem de anlayışını ve desteğini hiç esirgemeyen sevgili nişanlım Özlem'e minnet duyuyorum. Onların varlığıyla kendimi şanslı hissediyorum.

Ayrıca yüksek lisans öğrenimim sırasında; çalıştığım firmalardaki yöneticilerime verdikleri destek ve izin için teşekkür ederim.

Son olarak yüksek lisans çalışmamda verdiği maddi destekten dolayı TÜBİTAK'a teşekkür ederim.

Haziran 2010

Alper Özcan

Bilgisayar ve İşletme Mühendisi

İÇİNDEKİLER

Sayfa

ÖNSÖZ ..... **v**
İÇİNDEKİLER ..... **vii**
KISALTMALAR ..... **ix**
ÇİZELGE LİSTESİ ..... **xi**
ŞEKİL LİSTESİ ..... **xiii**
ÖZET ..... **xv**
SUMMARY ..... **xvii**
1. GİRİŞ **1**
1.1 Tezin Amacı 3
1.2 Tezin Yapısı 3
2. KONUYLA İLGİLİ ÇALIŞMALAR **5**
2.1 Geleneksel Web Öneri Sistemleri 5
2.2 Mobil Ortamdaki Öneri Sistemleri 10
2.3 Sosyal Ağ Bilgisini Kullanan Web Öneri Sistemleri 13
3. SOSYAL AĞ VERİSİ KULLANAN MOBİL ÖNERİ SİSTEMİ **15**
3.1 Veri Hazırlama 16
3.1.1 Sosyal ağ verisi hazırlama 16
3.1.2 Mekan ve kullanıcı verisi hazırlama 19
3.2 Rasgele Yürüyüş Algoritması Kullanılarak Topluluk Belirleme Yöntemi 20
3.3 Öneri Altyapı Mimarisi 27
3.4 Mobil Telefon Simülatörü 31
4. DENEYSEL SONUÇLAR **37**
4.1 Veri Hazırlama Sonuçları 37
4.2 Öneri Altyapısı Sonuçları 38
4.2.1 Kesinlik, anma ve f-ölçütü sonuçları 39
4.2.2 Rassal öneri modeliyle karşılaştırma 42
5. SONUÇ VE İLERİ ÇALIŞMALAR **45**
KAYNAKLAR **47**
EKLER **51**

KISALTMALAR

DS	**:** Data Set
DSR	**:** Data Set Random
GAB	**:** Group Asynchronous Browsing
GPS	**:** Global Positioning System
HTTP	**:** Hyper Text Transfer Protocol
JDK	**:** Java Development Kit
J2ME	**:** Java 2 Micro Edition
PDA	**:** Personal Digital Assistant
SOAP	**:** Simple Object Access Protocol
WSDL	**:** Web Services Description Language
XML	**:** Extensible Markup Language
3G	: Third Generation

ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1 : Ortam farkındalığı yöntemini kullanan mobil öneri sistemleri. 12
Çizelge 3.1 : Komşuluk Matrisi A. 22
Çizelge 3.2 : Derece Tablosu d. 23
Çizelge 3.3 : Geçiş Matrisi P t = 0 için.....23
Çizelge 3.4 : Geçiş Matrisi P t = 1 için.....24
Çizelge 4.1 : Tasarlanan Öneri Modeli Sonuçları (weight1 = 0.5, weight2 = 0.5)....40
Çizelge 4.2 : Tasarlanan Öneri Modeli Sonuçları (weight1 = 0.3, weight2 = 0.7)....41
Çizelge 4.3 : Tasarlanan Öneri Modeli Sonuçları (weight1 = 0.7, weight2 = 0.3)....41
Çizelge 4.4 : Tasarlanan Öneri Modeli Sonuçları (weight1 = 0.5, weight2 = 0.5)....42
Çizelge 4.5 : Rassal Kullanıcı Modeli Sonuçları. 43

ŞEKİL LİSTESİ

Sayfa

Şekil 3.1 : Örnek konuşma kayıtları. 18
Şekil 3.2 : Ayrıştırıcının kullanıldığı Web sayfası. 19
Şekil 3.3 : t=0 için G çizge yapısı. 22
Şekil 3.4 : Geçiş Çizgesi t=1 için. 24
Şekil 3.5 : Geçiş Çizgesi t=3 için. 25
Şekil 3.6 : Öneri altyapı mimarisi. 28
Şekil 3.7 : Kullanıcı-Ürün ve Ürün-Ürün matrisi. 29
Şekil 3.8 : İstemci-Sunucu mimarisi. 32
Şekil 3.9 : Kullanıcı ana giriş ekranı ve öneri görme ekranı. 33
Şekil 3.10 : Önerilen mekanları oylama ekranı....34
Şekil A.1 : Sınıf Diyagramı 52

SOSYAL AĞ VERİSİ KULLANILARAK MOBİL TELEFONLAR İÇİN ÖNERİ ALTYAPISI TASARLANMASI

ÖZET

Günümüzde artan mobil geniş bant teknolojisi ve üçüncü generasyon'a (3G) geçiş ile birlikte mobil telefonlar sesli konuşma kullanımlarının yanında televizyon izleme, müzik indirme, internete bağlanma gibi birçok yeni özellikleriyle de kullanılmaya başlanmıştır. Kısa zaman içinde daha da artacak veri aktarım hızları ile zaten pek çok konuda bilgisayarlarımızı yakalamaya başlamış akıllı telefon kullanımının artışı mobil telefonları hayatımızın vazgeçilmez parçalarından biri yapmaya başlayacaktır. Büyük olasılıkla kısa süre içinde adı da telefon olmaktan çıkarak, sesli iletişim bu cihazların pek çok özelliğinden biri olarak kalacaktır.

Mobil ortam ağları artan veri aktarım hızı ile birlikte haber, reklam, eğitim, alışveriş, dosya paylaşım gibi birçok hizmetin ucuz ve kolay sunulabildiği bir ortam halini almıştır. Hizmet sunmanın kolaylaştığı bu ortamda artan küresel rekabet mobil servis operatörlerinin, bireylere ya da gruplara özel olarak farklılaşan, etkili çözümler üretmesini gerektirmiştir. Bu açıdan müşteri memnuniyeti, eğilimleri ve kullanım alışkanlıkları baz alınarak, dinamik içerik içeren yapılar oluşturulmaktadır. Bu noktada mobil telefonlar için öneri sistemleri hem içeriksel hem de istenilen içeriğe daha kolay ve çabuk erişmeyi sağlayan çözümler üretmeye çalışmaktadır.

Genel olarak öneri sistemleri, kullanıcıların çeşitli alanlarda karar vermelerini kolaylaştıran yararlı bilgiler ya da ilgilenebilecekleri ürünleri, içeriği sunan öneriler üretmektedir. Önerilerin oluşturulmasında kullanıcı profilleri sıkça kullanılmaktadır. Mobil telefonlar için öneri sistemlerinde, buna ek olarak konum ve ortam bilgilerinin de kullanılması bu sistemlerin doğruluğunu arttırmaktadır. Ancak mobil telefonların ekran boyutlarının sınırlı olması ve mobil ağın sinyal gücünün çeşitli yerlerde farklılık göstermesi mobil telefon öneri sistemlerini olumsuz etkileyen faktörler olarak ortaya çıkmaktadır. Ayrıca mevcut mobil telefon öneri sistemleri genel olarak incelendiğinde sistemlerin kullanıcıların geri beslemesine ihtiyaç duyduğu görülmektedir. Bu noktada mobil telefon öneri sistemlerinde arkadaşlık bilgisinin ve arkadaş gruplarının tercih bilgisinin kullanılması doğruluğu ve performansı arttırıcı önemli bir iyileştirme olarak karşımıza çıkmaktadır. Bu nedenle bu tez çalışmasında arkadaşlık tercih bilgisinin öneri sistemine eklenmesi için kişilerin konuşma bilgisinden oluşan sosyal ağ bilgisinden yararlanılmıştır.

Bu tez çalışmasında, öncelikle mobil servis operatörleri için öneri sistemi altyapısı geliştirilmiştir. Geliştirilen altyapı mobil telefon kullanıcılarına istekleri doğrultusunda mekan önerileri üretmek üzere sınanmıştır. Bu altyapı dört aşamadan oluşmaktadır. Birinci aşamada kullanıcı mobil telefon arayüzü ile öneri isteğinde bulunmakta, ikinci aşamada kullanıcının konumu Küresel Yer Belirleme Sistemi (GPS) ile belirlenmekte, üçüncü aşamada içerik filtreleme, işbirlikçi filtreleme ve sosyal ağ analizi yöntemleri kullanılarak mekan önerileri üretilmekte, son aşamada ise kullanıcılara önerilen mekanlar çeşitli interaktif kanallar kullanılarak sunulmaktadır. Geliştirilen altyapı mimarisinin kullanımını modellemek için mobil telefon simülatörü tasarlanmış ve kullanıcılara mobil telefon arayüzü aracılığı ile mekan önerileri sunulmuştur. Öneri sisteminin kullandığı sosyal ağ verisi ülkemizde hizmet veren bir Mobil Servis Operatöründen elde edilmiş kişilerin gerçek konuşma verisinden oluşan veri kümesidir. Önerilen yöntemimize ait sonuçlar ve sonuçların genel geçerliliği çeşitli mekanlardan oluşan veri kümesi üzerinde sınanmıştır.

Önerdiğimiz yöntemin doğruluğu Kesinlik, Anma, F-Ölçütü gibi metriklerle ortaya konmuştur. Sonuçlara baktığımızda; sosyal ağ verisi kullanan öneri sistemi, kullanıcılara sunduğu mekan önerilerinde doğru sonuçlar üretmektedir. Elde edilen sonuç, birbirleriyle sık konuşan insanların benzer mekanlara gitme eğilimi gösterdiğini ve sık konuştuğu insanların gittiği mekanlarla ilgili önerileri alma eğiliminde olduğunu göstermektedir.

A RECOMMENDATION FRAMEWORK FOR MOBILE PHONES BASED ON SOCIAL NETWORK DATA

SUMMARY

Nowadays, with the growing mobile broadband technology and transition to third generation (3G), mobile phones started to be used not just to make voice calls but also for many other new features such as watching television, downloading music, and connecting to the Internet. With further and rapid increase of data transfer rates in near future, smart mobile phones will be one of the indispensables devices of our life as they are already preferred over computers today. Perhaps, very soon these devices will not be named as telephone and voice communication will be just one of the many features of them.

With increasing data transfer rates of mobile media network, services such as news, advertising, education, shopping, and file-sharing has become very cheap and easy to provide. Due to the increasing global competition caused by easy serving utilities of mobile media network, a distinctive, effective solutions for individuals and groups have to be produced. From this point of view, structures are built by dynamic contents based on customer care, propensity and usage habits.

In general, the recommendation systems generate useful information in order to support users when they make decisions or offer products they might like. The user profiles are frequently used to create recommendations. In addition, the usage of location and context information in mobile recommendation systems increases the accuracy of results. However, the limited screen size of mobile phones and the variation of mobile network signal strength in environment are two main negative factors affecting the system. When examined in details, it is also obvious that mobile recommendation systems need user feedback frequently. At this point, using the information of friendship and preferences of friends in mobile recommendation systems improve the accuracy and performance. Therefore, the social network information consisting of telephone calls is used to study in order to integrate the preferences of friends.

In this thesis, firstly a mobile recommendation system framework is developed for mobile service operators. This framework is tested to recommend places for mobile telephone users according to their preferences. This framework consists of four stages. In the first stage, the user requests for recommendation using mobile phone interface. Then at the second step, the user's location is determined by Global Positioning System (GPS). Using content and collaborative filtering with social network analysis methods, place recommendations are produced in the third stage. In the last step, the recommended places are returned using various mobile interactive channels. To model this recommendation framework architecture, a mobile phone simulator is designed and generated place recommendations are presented through a mobile phone interface. The social network data, which consists of real call data of users, are obtained from a mobile service operator in our country. The general validity of the proposed system is tested with a dataset consisting of various places.

The validity of the proposed method is measured in various metrics such as Precision, Recall and F-Measure. The results show that, the proposed system using social network data provides accurate results in place recommendations. The obtained results showed that people talking with each other tend to go similar places often and also tend to get information about the places where their friends visit.

1. GİRİŞ

Mobil Servis Operatörleri insanlara sınırların olmadığı bir dünya sunarken; eğitim, haber ve eğlence gibi çeşitli alanları kapsayan bilgi verme hizmetlerini bizlere sunmaktadır. Artan mobil geniş bant teknolojisi ve üçüncü generasyon'a geçiş ile birlikte sunulan bilgiye ve hizmetlere istenilen zamanda ve yerde ulaşmak gittikçe önem kazanmaktadır. Mobil ortamlarda sunulan hizmet ve kullanıcı sayılarının hızla artması, ancak mobil telefonların teknik özelliklerinin kısıtlı olması, yeni gereksinimler ortaya çıkarmıştır.

Günlük hayatımızda herhangi bir hizmeti ya da ürünü kullanma kararı verirken, çeşitli insanlardan gelen sözlü fikirler, pazarlama araştırmaları, gazete yazıları gibi öneri kaynakları tercihlerimizi şekillendirmektedir. Yaşamakta olduğumuz bilgi çağında üretilen bilgi ve içerik miktarı hızlı bir şekilde artmaktadır. Bu durum kullanıcıların ihtiyaçlarına uygun bilgi ve içerikleri bulan yöntemlerin geliştirilmesini gerekli kılmış ve öneri sistemlerinin önem kazanmasını sağlamıştır. Öneri sistemleri sayesinde kişilerin doğrudan ya da dolaylı tercih bilgileri toplanarak öneri modelleri oluşturulmakta ve kullanıcı tercihleri üzerinden öngörme ve öneri yapılmasına çalışılmaktadır. Öneri sistemleri ile ilgili akademik çalışmalara bakıldığında Web kullanıcıları için birçok çalışma bulunurken [1] mobil telefon kullanıcıları için çalışmalar azınlıktadır.

Mobil ortamdaki mevcut çalışmalar incelendiğinde mobil telefonların kısıtlı kullanıcı ara yüzleri ve mobil ağın sinyal gücünün çeşitli yerlerde farklılık göstermesi mobil telefon öneri sistemlerini olumsuz etkileyen faktörler olarak ortaya çıkmaktadır. Bu durum sunulan içerik ve hizmetlerin kullanılmasını zorlaştırmaktadır. Ayrıca öneri sistemlerin, daha doğru öneriler üretebilmek için kullanıcıların geri bildirimlerine ihtiyaç duyduğu görülmektedir. Bu nedenle mobil ağda kişilerin profillerine ve tercihlerine uygun içerik ve hizmet önerme çalışmaları önem kazanmaktadır. Bu noktada mobil öneri sistemlerinde arkadaşlık bilgisinin ve arkadaş gruplarının tercih bilgisinin kullanılması doğruluğu ve başarımı arttırıcı önemli bir iyileştirme olarak karşımıza çıkmaktadır.

Bu nedenle bu tez çalışmasında arkadaşlık tercih bilgisinin öneri sistemine eklenmesi için kişilerin konuşma bilgisinden oluşan sosyal ağ bilgisinden yararlanılmıştır. Konuşma verisinden oluşan sosyal ağda her bir düğüm kullanıcıları temsil etmektedir. Bu ağda birbirleriyle konuşan iki kullanıcı arasında bir ayrıt oluşturulmuş ve konuşma süresiyle orantılı olarak ağırlıklandırılmıştır. Birbirleriyle daha sık konuşmuş kullanıcıları belirlemek üzere rasgele yürüyüş algoritmasıyla [2] topluluklar bulunarak demetlenmiştir.

Sosyal ağ teorisine göre, sosyal topluluklar, her bir düğümün en az birkaç düğümle yakın ilişkide olduğu düğüm kümeleridir. Bu çalışmada sosyal topluluk yerine, sosyal gruplar da kullanılmıştır. Bir topluluk içindeki düğümler arasındaki ilişkiler daha yoğunken, farklı topluluklardaki düğümler arasındaki ilişkiler daha seyrektir [3].

Topluluk belirleme problemi arkadaşlık ağları [4], protein ağları [5], internet topolojisi [6], elektronik posta ağları [7], film ağları [8], alıntılama ağları [9] gibi birçok farklı araştırma alanını ilgilendirmektedir. Tez çalışmasında önerilen modelde, "*Birbirleriyle sık konuşan kişiler, günlük hayatlarında da birlikte zaman geçirme eğilimi gösterirler*" varsayımı ortaya konulmuştur. Bu varsayımı desteklemek için sosyal ağ verisi olarak ülkemizde hizmet veren bir mobil servis sağlayıcısından elde edilmiş kişilerin gerçek konuşma verisinden oluşan veri kümesinden yararlanılmış ve birbirleriyle sık konuşan topluluklar rasgele yürüyüş algoritması ile bulunmuştur. Kişisel gizlilik ilkesi doğrultusunda konuşma verisindeki kişilerin kimlik bilgileri veri kümesi içinde şifrelenmiş olarak tutulmaktadır. Çalışmada kullanılan mekan verisi ve bu mekanları ziyaret etmiş kullanıcılardan oluşan mekan tercih bilgileri geliştirilen ayrıştırıcı ile oluşturulmuştur. Veri hazırlama yöntemleri ve mekan önerileri üretmek üzere yapılan işlemler Bölüm 3'te ayrıntılı olarak anlatılmaktadır. Ardından mobil servis sağlayıcıları için öneri sistemi altyapısı geliştirilmiştir.

Geliştirilen altyapı dört aşamadan oluşmaktadır. Birinci aşamada kullanıcı mobil telefon arayüzü ile öneri isteğinde bulunmakta, ikinci aşamada kullanıcının konumu Küresel Yer Belirleme Sistemi (GPS) ile belirlenmekte, üçüncü aşamada içerik filtreleme, işbirlikçi filtreleme ve sosyal ağ analizi yöntemleri kullanılarak mekan önerileri üretilmekte, son aşamada ise kullanıcılara önerilen mekanlar çeşitli mobil interaktif kanalları kullanılarak sunulmaktadır. Geliştirilen altyapı mimarisinin kullanımını modellemek için mobil telefon simülatörü tasarlanmış ve kullanıcılara mobil telefon arayüzü aracılığı ile mekan önerileri sunulmuştur.

1.1 Tezin Amacı

Bu tezin amacı, mobil servis sağlayıcı operatörleri için sadece ortam ya da konum bilgisinin kullanıldığı öneri sistemlerine göre; hem konum ve ortam bilgisinin hem de oluşturulan sosyal ağ analizi bilgisinin kullanıldığı daha doğru sonuç veren bir öneri sistemi için altyapı oluşturulmasıdır. Oluşturulan öneri sisteminden, sosyal ağ yapısı kullanılarak kullanıcı geri bildirimine ihtiyaç duymayan ve öneri arama uzayını küçülten bunun yanında işbirlikçi filtreleme, içerik filtreleme ve sosyal ağ analizi yöntemlerinin birleştirildiği öneri altyapısı mimarisinin oluşturulması amaçlanmaktadır.

1.2 Tezin Yapısı

Tez çalışması beş bölümden oluşmaktadır. Bu çalışmada, konuşma verisinden oluşan sosyal ağ analizi bilgisinin, mobil öneri sistemlerinde nasıl uygulanabileceğini oluşturduğumuz mobil öneri sistemiyle tanıtmaktayız. Mobil ortamdaki öneri sistemleri ve sosyal ağ bilgisini kullanan öneri sistemlerine ilişkin çalışmalar 2. bölümde sunulmaktadır. Öneri sistemimize ait temel bileşenler ve bunların oluşturulmasında uygulanan yöntem ve kabuller 3. bölümde tanıtılmaktadır. 4. bölümde oluşturulan mekan veri kümesi üzerinden elde edilen deneysel sonuçlar sunulmaktadır. Son olarak, 5. bölümde sonuçlar ve ileriki çalışmalara ait yorumlar belirtilmektedir.

2. KONUYLA İLGİLİ ÇALIŞMALAR

Bu bölümde tez çalışmasının genel yapısını oluşturan, geleneksel öneri sistemlerinde kullanılan işbirlikçi filtreleme, içerik tabanlı filtreleme, sosyal ağ analizi yöntemleri ve mobil öneri sistemlerinde kullanılan ortam farkındalığı yöntemi genel olarak incelenmiştir.

2.1 Geleneksel Web Öneri Sistemleri

Öneri sistemleri, kullanıcıların daha önceki davranışlarını ve alışkanlıklarını inceleyerek, bir kullanıcıya uygun ürünler önermeyi hedefleyen uygulamalardır. Internet ortamında, bu ürünler, elektronik ticaret sitelerinde yer alan ürünler olabileceği gibi bir Web sitesindeki sayfalar da olabilir. Bu sistemler, kullanıcıların beğenilerine uygun ürünler önermek, kullanıcıya bir Web sitesinde yol göstermek, Web sitesinin yapısını iyileştirmek, kişiselleştirme veya bir Web sitesinin performansını artırma gibi çeşitli amaçlar için kullanılabilir. Literatürde öneri sistemleri için kullanılan çeşitli yöntemler mevcuttur. En sık uygulanan yöntemler olarak, işbirlikçi filtreleme [10], ilişkilendirme kuralları [11], Markov modelleri [12], demetleme yöntemleri [13], sıralı dizilerin [14] incelenmesi söylenebilir. Öneri sistemleri için ilk uygulanan ve bu nedenle de çok yaygın olarak kullanılan yöntemlerden biri işbirlikçi filtreleme yöntemidir.

Genel olarak geleneksel Web öneri sistemlerinde, belirli bir kullanıcının bir ürün hakkındaki tercih bilgisini tahmin etmek için işbirlikçi filtreleme yöntemi kullanılmaktadır. “Benzer kullanıcılar benzer ürünleri beğenmiştir” öngörüsü işbirlikçi filtreleme yönteminin arkasında yatan genel fikirdir. İşbirlikçi filtreleme yöntemi bellek tabanlı işbirlikçi filtreleme ve model tabanlı işbirlikçi filtreleme olmak üzere iki ana kısma ayrılmaktadır [15].

Bellek tabanlı işbirlikçi filtreleme diğer bir adıyla kullanıcı tabanlı işbirlikçi filtreleme yöntemini kullanan algoritmalar, kullanıcıların tercih örüntülerine göre kullanıcılar arasındaki korelasyon ve ilişkiyi hesaplamaya çalışmaktadır. Korelasyon değeri öneri yapılırken farklı kullanıcı grupları arasında dinamik olarak hesaplanmaktadır. Gerçek zamanlı öneri sistemlerinde bu dinamik hesaplama büyük bir bellek ve işlem yüküne neden olmaktadır. Bu nedenle, bu tarz öneri sistemleri büyük veri kümeleri ve gerçek zamanlı uygulamalar için uygun değildir. Temel olarak bellek tabanlı işbirlikçi filtreleme algoritmaları öneri isteğinde bulunan aktif kullanıcı ile veritabanındaki diğer kullanıcıların benzerliklerini bulmaya çalışmaktadır. Algoritmaların bellek tabanlı olarak tanımlanmasının nedeni algoritmaların kullanıcı veritabanının bütününde işlem yapmasıdır. Bütün kullanıcıların tercihleri, kullanıcıların ürünlere vermiş olduğu oylarla simgelenmektedir. Aktif kullanıcının oy değeri oyladığı ürünlerin oy değerinin ortalama değeri olmaktadır. Aktif kullanıcının tahmin edilen oyları diğer kullanıcı oylarının ağırlıklı toplamları eklenerek hesaplanmaktadır. Ağırlıklar aktif kullanıcı ile diğer kullanıcıların benzerlikleri tarafından belirlenmektedir bu nedenle kullanıcıların benzerlikleri arttıkça ağırlıklı toplama katkısı da artmaktadır [16].

Kullanıcının ortalama oyu Eşitlik **(2.1)**'deki gibi hesaplanmaktadır. I_i kullanıcı i'nin oyladığı ürünlerin kümesini, v_{ij} kullanıcı i'nin j ürününü oylama değerini simgelemektedir.

$$\bar{v}_i = \frac{1}{|I_i|} \sum_{j \in I_i} v_{i,j} \qquad \textbf{(2.1)}$$

Aktif kullanıcının tahmin edilen oyu Eşitlik **(2.2)**'deki gibi hesaplanmaktadır.

$$p_{a,j} = \bar{v}_a + k \sum_{i=1}^{n} w(a,i)(v_{i,j} - \bar{v}_i) \qquad \textbf{(2.2)}$$

Eşitlik **(2.2)**'de k normalize faktörünü, $w(a,i)$ kullanıcı i'nin aktif kullanıcıya katkısının ağırlığını göstermektedir. Ağırlık değerleri aktif kullanıcı ile veritabanındaki diğer kullanıcıların oyladıkları ortak ürünlerin karşılaştırılması ile bulunmaktadır.

Ağırlık değerleri; Eşitlik **(2.3)**'teki ortalama karesel farkın tersi [17], Eşitlik **(2.4)**'teki Pearson korelasyonu [18] ve Eşitlik **(2.5)**'teki vektör benzerliği [19] formülleri ile belirlenebilir.

$$w(a,i) = \frac{1}{(V_{i,j} - V_{a,j})^2} \tag{2.3}$$

Eşitlik **(2.3)**'de v_{ij} kullanıcı i'nin j ürününü oylama değerini simgelemekte, v_{aj} ise aktif kullanıcının j ürününü oylama değerini simgelemektedir.

$$w(a,i) = \frac{\sum_{j=1}^{n} (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_{j=1}^{n} (v_{a,j} - \bar{v}_a)^2 \sum_{j=1}^{n} (v_{i,j} - \bar{v}_i)^2}} \tag{2.4}$$

Yine aynı şekilde Eşitlik **(2.4)**'de v_{ij} kullanıcı i'nin j ürününü oylama değerini, v_{aj} aktif kullanıcının j ürününü oylama değerini göstermektedir.

$$w(a,i) = \sum_{j=1}^{n} \frac{v_{a,j}}{\sqrt{\sum_{k \in I_a} v_{a,k}^2}} \frac{v_{i,j}}{\sqrt{\sum_{k \in I_i} v_{i,k}^2}} \tag{2.5}$$

Eşitlik **(2.5)**'de I_i kullanıcı i'nin oyladığı ürünlerin kümesini, I_a aktif kullanıcı a'nın oyladığı ürünlerin kümesini göstermektedir.

Model tabanlı işbirlikçi filtreleme diğer bir adıyla ürün tabanlı işbirlikçi filtreleme büyük veri kümelerinde ölçeklenebilir olması nedeniyle güncel uygulamalarda yaygın olarak kullanılmaktadır. Model tabanlı algoritmalar kullanıcılar arasındaki benzerlikleri hesaplamak yerine tercih edilmiş ürünler arasındaki benzerlikleri hesaplamaktadır. Temel olarak model tabanlı işbirlikçi filtreleme algoritmaları öncelikle kullanıcıların tercihlerini derleyerek tanımlayıcı bir model oluşturup ardından modele uygun olarak öneri üretmektedirler. Olasılık bakış açısına göre model tabanlı işbirlikçi filtreleme yöntemi öneri yapılacak aktif kullanıcının profili ve kullanıcının önceki oylarına göre oyların beklenen değerini bulmak şeklinde tanımlanmış ve beklenen değer bulma formülü Eşitlik **(2.6)**'deki gibi verilmiştir.

$$P_{a,j} = E(v_{a,j}) = \sum_{i=0}^{m} \Pr(v_{a,j} = i \mid v_{a,k}, k \in I_a) \cdot i \qquad \textbf{(2.6)}$$

Breese ve diğ. (1998) tarafından tasarlanan demetleme modelinde aynı tercihleri paylaşan kullanıcılar aynı grupta sınıflandırılmaktadırlar. Belirli bir sınıftaki aktif kullanıcının oyları bağımsız kabul edilmekte, sınıfın ve oyların birleşik olasılığı Naive Bayes formülü ile Eşitlik **(2.7)**'deki gibi hesaplanmaktadır [20].

$$\Pr(C = c, v_1, ..., v_n) = \Pr(C = c) \prod_{i=1}^{n} \Pr(v_i \mid C = c) \qquad \textbf{(2.7)}$$

Buna ek olarak işbirlikçi filtreleme sistemlerinde, yeni bir ürün ya da yeni bir kullanıcı sisteme eklendiğinde yeni eklenen kullanıcı herhangi bir ürünü oylayıncaya kadar ya da yeni eklenen ürün herhangi bir kullanıcı tarafından oylanıncaya kadar sistemin kullanıcılar ve ürünler arasındaki benzerliği hesaplayabilecek bilgisi bulunmamaktadır. Bu problemin ismine en baştan başlatma (cold-start) problemi adı verilmektedir. Bu problemi çözmek için içerik tabanlı filtreleme yöntemleri önerilmiştir. İçerik tabanlı filtreleme yönteminde ürünlerin ve kullanıcı tercih bilgilerinin içeriğinden elde edilen özellik vektörleri kullanılmaktadır. Ürün ve kullanıcı tercih vektörleri arasındaki benzerlikler bulunarak önerilecek ürünler belirlenmektedir [21].

En baştan başlatma problemini ortadan kaldırmak için Fab öneri sistemi, işbirlikçi filtreleme yöntemi ile içerik tabanlı filtreleme yöntemini birleştirmiştir [22]. Fab öneri sistemi, işbirlikçi filtreleme ve içerik tabanlı filtrelemeyi birleştiren melez bir yöntem olup Web tabanlı bir öneri hizmetidir. Kullanıcı profilleri, her bir kullanıcının yüksek oy ile oyladığı sayfaların anahtar kelimelerinden oluşturulmaktadır. Bir sayfanın içeriği yüksek oy alan başka bir sayfanın içeriği ile eşleşiyorsa ya da benzer kullanıcılar tarafından bir sayfa yüksek oy almış ise bu sayfa öneri olarak kullanıcıya sunulmaktadır. Olumlu ve olumsuz oylamalar sonucunda oylamanın etkisini profile yansıtmak için kullanıcının profili güncellenmektedir. Fab sisteminin en önemli katkısı içerik tabanlı filtreleme ve işbirlikçi filtreleme yöntemlerinin özelliklerini birleştirmesi ve öneri sistemini kullanıcıların tercihlerine göre dinamik olarak güncel tutmasıdır.

Turnbull (1997) tarafından tasarlanan Firefly öneri sistemi başlangıçta müzik ve film önerileri için kullanılmakta olup günümüzde haber, kitap ve Web sayfası önerilerinde de kullanılmaktadır [23]. Kullanıcı benzerliklerine göre öneride bulunan sistem kullanıcı profilini giriş olarak almakta ve Pearson algoritmasını kullanarak öneride bulunmaktadır. Algoritma temel olarak belirli ürünleri beğenen ve beğenmeyen kullanıcılara göre kullanıcı profillerini oluşturmakta ardından kullanıcıların profillerine göre benzerliklerini karşılaştırarak öneride bulunmaktadır.

Wittenburg ve diğ. (1998) tarafından tasarlanan GAB öneri sistemi Web sayfası önerilerinde kullanılmaktadır [24]. GAB Öneri Sisteminde asenkron grup tarama yöntemi kullanılarak sık kullanan sayfalar belirlenip kullanıcıların etkin sayfa listesi hazırlanmakta ve hazırlanan etkin sayfa listesi kullanıcılara öneri olarak sunulmaktadır. Sistem kullanıcıların sık kullanılan sayfalarına erişebildiği için kişiye özel gizlilik sorunlarına neden olmaktadır. Gizlilik problemini gidermek için öneri sistemi kullanıcılara sık kullanılan sayfaları özel ya da genel olarak kaydedebilecekleri bir araç sunmaktadır.

2.2 Mobil Ortamdaki Öneri Sistemleri

Kullanıcıların tercihlerine uygun önerilerin bulunması için öneri sistemlerinde kişiselleştirme teknikleri üzerinde birçok çalışma yapılmaktadır. Ayrıca mobil cihazların kısıtlı kullanıcı arayüzleri ve iş akış mantıkları nedeniyle, kullanıcılara zamanında ve uygun önerilerde bulunmak için mobil telefon öneri sistemlerinin mobil telefonlarda kullanmaya uygun hale getirilmesi gerekmektedir. Mobil servis sağlayıcılar kullanıcılara fiziksel konumları ile ilişkili bilgi ve hizmet sunan konum tabanlı servisler sunarak kullanıcıların yararlandıkları mobil hizmetleri kişiselleştirmeyi amaçlamaktadır.

Genel olarak mobil ortamlardaki öneri sistemlerinin karakteristik özellikleri incelendiğinde, mobil öneri sistemleri kullanıcılar hareket halindeyken hizmet sağlamak için ortam farkındalığına (context-aware) sahip olmalıdır. Ortam farkındalığı yöntemi lokanta önerilerinde, tur rehberlerinde ve ticari öneri sistemlerinde kullanılmaktadır. Mobil ağdaki ortam tanımı şu şekildedir. Ortam bir mekanın, kişinin ya da nesnenin durumunu ayırt etmek için kullanılan bilgilerin bütünüdür. Bir öneri sistemi içerik ve hizmet önerilerinde bulunurken ortam bilgisini kullanıyorsa ortam tabanlı sistem olarak düşünülmektedir [25].

Mobil öneri sistemleri için zaman, coğrafi konum, kimlik, cihaz bilgisi gibi birçok ortam bilgisi kullanılabilir. Coğrafi konum kullanıcıya ait mobil cihazın bulunduğu konumu göstermekte, zaman bilgisi öneri isteğinin geldiği zaman bilgisini belirtmekte, kimlik bilgisi ise mobil cihazın o andaki kullanıcısını tanımlamaktadır. Dey ve Abowd (2000) kullanıcının anlık durumunu tanımlamak için etkinlik bilgisini de ortam bilgisi arasına eklemiştir [26]. Aktivite bilgisi kullanıcının öneri isteğinde bulunduğu andaki yaptığı işi tanımlamaktadır. Birçok akademik çalışma mobil ağda ortam bilgisini kullanarak kullanıcılara hizmet ve içerik önerilerinde bulunmaktadır. Çizelge **2.1**'de ortam tabanlı bazı mobil öneri sistemleri özetlenmiştir.

Tung ve Soo (2004)'nun tasarlamış olduğu kişiselleştirilmiş restoran öneri sistemi prototipinde, öneri yöntemi kullanıcı ile sistem arasındaki iletişime dayanmaktadır. Öneri üretilme aşamasında kullanıcıdan alınan kısıtlar kullanılmaktadır. Bu çalışmada uygulamanın geliştirildiği donanım olarak dijital kişisel yardımcı (PDA) belirlenmiş, kullanıcının konumunun belirlenmesi için de küresel konumlandırma sistemi (GPS) kullanılmıştır. Öneri yöntemi beş aşamadan oluşmaktadır. Birinci aşamada kullanıcıdan öneri isteği ve kullanıcının belirttiği kısıtlar alınmakta, ikinci aşamada kullanıcının ortam bilgisi belirlenmekte, üçüncü aşamada ortam bilgisine göre uygun bilgiler aranmakta, dördüncü aşamada kullanıcının kısıtları çözümlenmekte, son aşamada ise önerilen restoranlar kullanıcıya sunulmaktadır. Tasarlanan prototipin kullanılabilirliğini göstermek için örnek bir kullanıcıya restoran önerilerinde bulunulmuştur [27].

Park ve diğ. (2007)'nin çalışmasında ise mobil cihazlarda Bayes ağı modeli kullanılarak konum tabanlı gerçek zamanlı öneri sistemi tasarlanmıştır. Tasarlanan öneri sistemi ortam kaydı toplama modülü, öneri modülü ve eşleştirme modülü olmak üzere üç kısımdan oluşmaktadır. Kullanıcı profili ve ortam bilgisi mobil cihazdan toplanmakta, ardından Bayes ağının parametrelerini eğitmek için toplanan ortam bilgileri önişlemden geçmektedir. Kullanıcı yeni bir öneri isteğinde bulunduğunda, eğitilmiş Bayes ağı modelinin en yüksek olasılık değerli parametresi seçilmektedir. Ayrıca mobil cihazların ekran kısıtlarını aşmak için harita tabanlı bir arayüz oluşturulmuştur. Kullanıcıların kişisel özelliklerini modellemek için kullanıcıların isim, cinsiyet, yaş, doğum tarihi, kan grubu, aylık gelir, tercih edilen yemek gibi bilgileri toplanmaktadır. Önerilecek uygun bir restoran bulunması için Bayes ağının çıkarsama sonucuna ait restoran tercihleri hesaplanmaktadır. Bir restoran sınıf, fiyat ve atmosfer olmak üzere üç özelliğe sahiptir. Bir restoranın öneri puanı hesaplanırken bu üç özellik kullanılmaktadır. Sunulan öneri modelinin çalışmasının gösterilmesi için 50 restoran ve 4 kullanıcı için yedi günlük ortam bilgisi toplanarak kullanıcılara önerilerde bulunulmuştur [28].

Çizelge 2.1 : Ortam farkındalığı yöntemini kullanan mobil öneri sistemleri.

Yazarı ve Yılı	Öneri Tipi	Ortam Bilgisi
Cheverst ve diğ. 2000 [29]	Tur rehberi	Konum, zaman
Tewari ve diğ. 2002 [30]	Restoran	Konum
Yuan ve diğ. 2003 [31]	Reklam	Konum, zaman, içerik bilgisi
Pousman ve diğ. 2004 [32]	Zaman planlaması	Konum
Hayes ve diğ. 2004 [33]	Müzik	Kullanıcı tercihleri
Tung ve diğ. 2004 [27]	Restoran	Konum, zaman, hava durumu
Setten ve diğ. 2004 [34]	Tur rehberi	Konum, zaman, hava durumu

Bir diğer çalışma olan Ricci ve Nguyen (2005)'in yorum tabanlı mobil öneri sistemi çalışmasında ise interaktif olarak kullanıcıdan alınan yorumlarla kullanıcının ihtiyaçlarını ortaya çıkarıp öneride bulunan MobyRek isimli mobil öneri sistemi tasarlanmıştır. Sistem kullanıcıya aday ürünlerin sunulması aşamasında kullanıcıyla iletişime geçerek önerilen ürünler hakkında kullanıcıdan geri bildirimler ve yorumlar beklemektedir. Ardından kullanıcı yorumları sistem tarafından değerlendirilmekte ve kullanıcının tercih modeli ile birleştirilmektedir. Bu şekilde aşama aşama kullanıcıdan geri bildirim ve yorum alarak kullanıcıların tercihleri ve ihtiyaçları belirlenmekte ve kullanıcılara restoran önerilerinde bulunulmaktadır. Öneri süreci kullanıcının öneri isteğinde bulunmasıyla başlar, sonuç olarak önerilen mekanlardan birini seçmesiyle ya da hiçbirini seçmemesiyle tamamlanır. Öneri/yorum modunda çalışan öneri sistemi, kullanıcıların sistemin sorduğu soruları cevaplamasını gerektirdiği için öneri aşamasında sürekli kullanıcı ile iletişim halinde olmasını gerektirmektedir [35].

Son olarak Lee ve diğ. (2005)'nin mobil ağda ortam tabanlı öneri yöntemi çalışmasında, mobil ortam bilgisi olarak konum, zaman, kullanıcı kimlik bilgisi, cihaz bilgisi ve kullanıcının yaptığı anlık etkinlik gibi bilgiler kullanılarak tasarlanan sistem Kore'deki mobil servis sağlayıcısında uygulanmıştır. Kullanıcı kayıtları analiz edilerek sunulan modelin performansı ölçülmüştür. Koreli mobil servis sağlayıcısına kayıtlı 200 kullanıcı tarafından ortam tabanlı öneri yöntemi analiz edilmiş ve Mart 2005 – Mayıs 2005 tarihleri arasındaki haber, sinema ve restoran hizmetlerinin kullanım kayıtları incelenerek ortam tabanlı öneri yönteminin performansı ölçülmüştür. Sonuç olarak eğer kullanıcılar yapılan önerileri uygun bulmuşsa pozitif, uygun bulmamışsa negatif geri bildirim yaparak servis sağlayıcıya bildirmişlerdir. Pozitif geri bildirim değeri toplam öneri sayısına oranlanarak ortalama doğruluk değeri ölçülmüş ve 0.579 olarak belirlenmiştir [25].

2.3 Sosyal Ağ Bilgisini Kullanan Web Öneri Sistemleri

Günümüzde Friendster[1], Orkut[2] ve MySpace[3] gibi Web ortamında sosyal ağ tabanlı öneri sistemleri kitap, müzik, film gibi birçok alanda önerilerde bulunmaktadır. Tarveen ve Hill (2001) öneri sistemlerinin Web ortamındaki sosyal aktivite izlerinden beslenmesi gerektiği fikrini ortaya koymuştur ve bu yaklaşım sosyal veri madenciliği olarak adlandırılmıştır [36]. Liu ve Maes (2005)'in geliştirdiği tercih eşleştirmesi yöntemi ile kullanıcıların Web ortamında bıraktığı izlerden anlam çıkararak kişilerin günlük ilgi ve tercihleri tahmin edilmeye çalışılmıştır. Geliştirilen yöntemde kişilerin ilgi ve tercihlerinin tanımlandığı 100000 sosyal ağ profili toplanmıştır. Doğal dil işleme teknikleri ile Web sayfalarının yapısal analizleri yapılarak sosyal ve dinamik kaynaklardan çeşitli bilgiler toplanıp kullanıcı profilleri zenginleştirilmekte ve profiller arasındaki benzerlik örüntüleri analiz edilerek tercihlerin ve kullanıcıların birbirlerine bağlı olduğu tercih eşleştirme ağı oluşturulmuştur. Tercih eşleştirme yöntemi dört aşamada gerçekleşmektedir. Birinci aşamada sosyal ağ profillerinin çıkartılması, ikinci aşamada profilleri temsil eden sözcük grupları ve anahtar kelimelerin tercih tanımlayıcılarından oluşan ontolojiye eşleştirilmesi, üçüncü aşamada profil verilerinin içeriği hakkındaki verilerin (metadata) profillere eklenmesi ve son aşamada da makine öğrenme tekniklerinin uygulanarak tercih tanımlayıcılar arasındaki anlamsal ilişkilerin öğrenilmesi gerçekleştirilmektedir. Tercih eşleştirme ağında her bir düğüm tercih tanımlayıcılarını düğümler arasındaki bağlarda anlamsal ilişkiyi temsil etmektedir. Çalışmada yapılan değerlendirme sonucunda tercih eşleştirme ağına bağlı olarak yapılan öneriler kişilerin tercihlerine uygun ve anlaşılabilir olduğu gösterilmiştir [37].

[1] http://www.friendster.com/
[2] http://www.orkut.com/
[3] http://www.myspace.com/

Golbeck (2006) tarafından geliştirilen FilmTrust isimli sosyal ağ tabanlı Web öneri sisteminde, sosyal ağ analizi ve işbirlikçi filtreleme yöntemleri kullanılarak film önerilerinde bulunulmuştur. Geliştirilen yöntemde kullanıcılar arkadaş olarak eklediği kişilere güven oyu vererek eklemiş oldukları arkadaşlarının tercih ettiği filmlere güvendiğini belirtmektedir. Ayrıca, kullanıcılar istedikleri filmi oylayabilmekte ve yorum yazabilmektedir. Bir kullanıcı ile film arasındaki öneri puanı bulunurken, ilgili filmi oylayan kullanıcılar ile ilgili kullanıcı arasındaki güven değeri öneri puanını etkilemektedir. Sonuç olarak öneri puanının bulunmasında, güven değerinin kullanıldığı önerilerin doğruluğu, güven değerinin kullanılmadığı önerilere göre daha büyük çıkmıştır. Sunulan bu modelde sosyal ağlar ile kişilerin birbirlerine vermiş olduğu güven değeri birleştirilerek film önerilerinde bulunulmuştur [38].

3. SOSYAL AĞ VERİSİ KULLANAN MOBİL ÖNERİ SİSTEMİ

Tez çalışmasında tasarlanan mobil öneri sisteminde işbirlikçi filtreleme, içerik filtreleme ve sosyal ağ analizi yöntemlerinin birleştirildiği öneri altyapı mimarisi oluşturulmuş ardından mobil servis sağlayıcı operatörleri için ortam, konum ve sosyal ağ analizi bilgisinin kullanıldığı öneri sistemi gerçeklenmiştir. Geliştirilen sistemin kullanımını modellemek için mobil telefon simülatörü tasarlanmış ve kullanıcılara mobil telefon arayüzü aracılığı ile mekan önerileri sunulmuştur. Öneri sistemi temel olarak üç ana bileşenden oluşmaktadır: Sosyal Ağ yapısı, Veri altyapısı ve Öneri yapısı.

İlk bileşen olan sosyal ağ yapısı; ülkemizde hizmet veren bir mobil servis operatöründen elde edilmiş gerçek konuşma verisinden oluşmaktadır. Konuşma verisindeki kullanıcı bilgileri kişisel gizlilik ilkesi gereği şifreli olarak tutulmuştur. Bu veriden düğümleri kullanıcılar olan bir çizge oluşturulmuş, birbiriyle telefon görüşmesi yapan kişiler için çizgede ilgili düğümler arasına bir ayrıt eklenmiştir. Bu kişilerin konuşma süreleri normalize edilerek ayrıt ağırlığı belirlenmiştir. Birbirleriyle daha çok konuşan kişilerin belirlenip demetlenmesi için rasgele yürüyüş algoritması kullanılarak sık konuşmuş kişiler topluluklara ayrılmıştır.

İkinci bileşen olan veri altyapısı, restoran veritabanı ve kullanıcı veritabanından oluşmaktadır. Restoran veritabanı, İstanbul'un çeşitli semtlerinde bulunan restoran tarzı mekanlardan elde edilmiştir. Kullanıcı veritabanı ise bu mekanları ziyaret etmiş ve mekanlar hakkında yorum yazmış kullanıcı bilgilerinden oluşmaktadır. Her iki veritabanını oluşturmak için öncelikle bir ayrıştırıcı (parser) geliştirilmiş ardından Web ortamında mekanların ve bu mekanlara giden kullanıcı bilgilerinin bulunduğu bir Web sitesindeki[4] veriler ayrıştırıcı ile toplanarak veritabanına kayıt edilmiştir.

[4] http://www.mekanist.net/

Son bileşen olan öneri yapısında; ilk iki bileşende oluşturulan kullanıcı ve restoran veritabanı ile sosyal ağ yapısı kullanılmaktadır. Geliştirilen öneri modeli dört aşamadan oluşmaktadır. Birinci aşamada kullanıcı mobil telefon arayüzü ile öneri isteğinde bulunmakta, ikinci aşamada kullanıcının konumu Küresel Yer Belirleme Sistemi (GPS) ile belirlenmekte, üçüncü aşamada içerik filtreleme, işbirlikçi filtreleme ve sosyal ağ analizi yöntemleri kullanılarak mekan önerileri üretilmekte, son aşamada ise kullanıcılara önerilen mekanlar çeşitli mobil interaktif kanalları kullanılarak sunulmaktadır. Geliştirilen altyapı mimarisinin kullanımını modellemek için mobil telefon simülatörü tasarlanmış ve kullanıcılara mobil telefon arayüzü aracılığı ile mekan önerileri sunulmuştur. Bu bölümde, yukarıda bahsedilen bileşenlerin oluşturulmasına ait uygulanan yöntem ve kabuller tanıtılmaktadır.

3.1 Veri Hazırlama

Bu tez çalışmasında, kullanıcılara ait gerçek konuşma verisi ve Web ortamından ayrıştırıcı ile elde edilmiş mekan verileri kullanılacaktır. Bu kaynak verileri işlenmemiş halde alınıp, sonraki aşamalarda kullanılabilecek biçime getirilmesi aşağıdaki Bölüm 3.1.1 ve Bölüm 3.1.2'de anlatılmaktadır.

3.1.1 Sosyal ağ verisi hazırlama

Çalışmada kullanılan sosyal ağ verisi ülkemizde hizmet veren bir mobil servis operatöründen elde edilmiş 340.176 farklı kullanıcıya ait 2008 Ocak ayında gerçekleşen ülke genelindeki gerçek konuşma verisinden oluşmaktadır. Servis operatörünün kayıtlarına göre, kullanıcılardan 14.128'i sadece aranmış, ancak herhangi birisini aramamıştır. Mevcut kayıtlar sadece arama kayıtlarından oluşmaktadır, tekrar olacağı sebebiyle aranma kayıtları alınmamıştır. Konuşma verisindeki kullanıcı bilgileri kişisel gizlilik ilkesi gereği şifreli olarak tutulmuştur. Dosya formatındaki konuşma verileri tez çalışmasında veritabanı sunucusu olarak kullanılan MySQL Server 5.1 ortamında yaratılan *konusma_cizgesi* tablosuna kaydedilmiştir.

Yaratılan konusma_cizgesi tablosunda 20.934.983 satır kayıt bulunmaktadır. Kayıtların içeriği incelendiğinde konuşma verisi Şekil **3.1**'de görüldüğü gibi arayan kişinin numarası, aranan kişinin numarası, aramanın nereden yapıldığı, aramanın hangi tarihte yapıldığı, aramanın kaç saniye sürdüğü, arama sonucunda verilen ödül süre miktarı, aramayı yapan kişinin segment bilgisi, abonelik aktivasyon tarihi, abonelik tarife bilgisi, kişinin eğitim bilgisi gibi çeşitli bilgilerden oluşmaktadır. Kayıtlardaki veri alanlarından bu çalışmada sadece İstanbul içinde yapılan arama kayıtları ayıklanarak arayan kişinin numarası, aranan kişinin numarası, arama süresi verileri kullanılmıştır. İstanbul içinde oluşan arama kayıtları ayrı bir geçici tabloya alınmıştır. Eğer iki kullanıcı birden fazla konuşma yapmış ise tabloda birden fazla konuşma kaydı bulunmaktadır. Bu kayıtları bire indirmek için, belirli iki kullanıcı arasında birden fazla olan konuşma kayıtlarının süreleri toplanarak her kullanıcı çifti arasında tek kayıt oluşturulmuştur. Ayıklama işlemi sonucunda İstanbul içinde birbirleri ile konuşmuş 129.919 tane aboneye ait konuşma kaydı elde edilmiştir. Sosyal ağ analizi yöntemi ile toplulukların bulunabilmesi için konuşma kayıtlarının konuşma çizgesinde temsil edilmesi gerekmektedir. Ayıklanan konuşma kayıtlarının çizge üzerinde temsil edilebilmesi ve rasgele yürüyüş algoritması ile bulunan topluluk yapılarının anlamlı çıkması için veri kümesi son bir ayıklama işleminden daha geçmektedir. Bu ayıklama işlemi için öncelikle en az 5 farklı aboneyi aramış olan numaralar bulunmaktadır. Ardından bulunan numaraların aramış olduğu abonelerden, en az 3 farklı aboneyi aramış olan kişiler bulunarak algoritmanın işlem yapacağı veri kümesine dahil edilmektedir. Bunun nedeni algoritmanın işlem yapacağı çizgedeki düğümlerin bağlı olmasını sağlamak ve ayrık çizge yapılarının oluşmasını engellemektir. Bu şekilde yapılan son ayıklama işleminden sonra 3.949 kullanıcının bulunduğu bir veri kümesi elde edilmiştir.

caller_id	callee_id	call_group_tp	call_date	time_band_id	termination_id	traffic_group_id	duration	free_duration	cell_id	call_count	bonus
9575231	14	VOICE	2008-01-04 17:29:13	18	1210	1	15	0	22590	1	NULL
9575231	14	VOICE	2008-01-04 17:31:39	18	1210	1	15	0	16354	1	NULL
8341232	14	VOICE	2008-01-05 10:18:26	11	1210	1	22	0	2273	1	NULL
8341232	14	VOICE	2008-01-05 10:32:52	11	1210	1	13	0	2273	1	NULL
8341232	14	VOICE	2008-01-05 10:28:36	11	1210	1	9	0	2273	1	NULL
8341232	14	VOICE	2008-01-05 10:20:50	11	1210	1	30	0	2273	1	NULL
9575231	14	VOICE	2008-01-06 16:11:05	17	1210	1	42	0	26562	1	NULL
9575231	14	VOICE	2008-01-12 09:41:24	10	1210	1	31	36	26562	1	3826
9575231	14	VOICE	2008-01-12 09:41:24	10	1210	1	31	0	26562	1	NULL
9575231	14	VOICE	2008-01-13 11:42:11	12	1210	1	15	0	26562	1	NULL
9575231	14	VOICE	2008-01-13 11:42:11	12	1210	1	15	18	26562	1	3826
9575231	14	VOICE	2008-01-20 15:29:26	16	1210	1	7	12	26562	1	3826
9575231	14	VOICE	2008-01-20 15:29:26	16	1210	1	7	0	26562	1	NULL
8811858	14	VOICE	2008-01-23 15:11:59	16	1210	1	99	0	7933	1	NULL
8811858	14	VOICE	2008-01-23 15:11:59	16	1210	1	99	102	7933	1	3863
9575231	14	VOICE	2008-01-26 11:30:35	12	1210	1	33	36	26561	1	3826
9575231	14	VOICE	2008-01-26 11:30:35	12	1210	1	33	0	26561	1	NULL
8811858	14	VOICE	2008-01-29 20:37:03	21	1210	1	90	0	2661	1	NULL
8811858	14	VOICE	2008-01-29 20:37:03	21	1210	1	90	90	2661	1	3863
11811781	58	VOICE	2008-01-01 01:45:31	2	1210	1	16	0	12171	1	NULL
11811781	58	VOICE	2008-01-01 01:45:31	2	1210	1	16	18	12171	1	3863
7123720	58	VOICE	2008-01-02 15:58:56	16	1210	1	4	0	15153	1	NULL
7123720	58	VOICE	2008-01-02 16:23:39	17	1210	1	20	24	14507	1	3861
7123720	58	VOICE	2008-01-02 16:30:36	17	1210	1	23	0	14507	1	NULL
11811781	58	VOICE	2008-01-02 16:48:51	17	1210	1	137	138	26015	1	3863
7123720	58	VOICE	2008-01-02 13:49:05	14	1210	1	42	42	17606	1	3861
7123720	58	VOICE	2008-01-02 15:48:07	16	1210	1	12	0	1131	1	NULL
7123720	58	VOICE	2008-01-02 16:30:36	17	1210	1	23	24	14507	1	3861
7123720	58	VOICE	2008-01-02 16:41:44	17	1210	1	67	0	32674	1	NULL
7123720	58	VOICE	2008-01-02 13:49:05	14	1210	1	42	0	17606	1	NULL
11811781	58	VOICE	2008-01-02 16:48:51	17	1210	1	137	0	26015	1	NULL
7123720	58	VOICE	2008-01-02 16:23:39	17	1210	1	20	0	14507	1	NULL

Query aborted. 127170 rows fetched so far in 0.8853s (0.0279s)

Edit | Apply Changes | Discard Changes | First | Last

Şekil 3.1 : Örnek konuşma kayıtları.

3.949 kullanıcıdan oluşan son veri kümesi rasgele yürüyüş algoritması kullanılarak sık konuşmuş kişiler 271 adet topluluğa ayrılmıştır. Ardından çalışmada kullanılan restoranlara giden 551 kullanıcı ile eşlemek için toplam 551 kullanıcıdan oluşan 10 topluluk belirlenmiş ve bu veri kümesi Bölüm 4'teki deneysel çalışmada kullanılmıştır. Restoranlara giden 551 kullanıcı elde edilebildiği için belirlenen 10 topluluk 551 kullanıcıdan oluşmaktadır. Toplulukları belirlemek için algoritmanın bulduğu bütün topluluk yapıları incelenmiş, düğüm sayıları az olan topluluk yapıları ayıklanmıştır. Sonuç olarak belirlenen on topluluktan, dokuz tanesi 55 düğümden, bir topluluk ise 56 düğümden oluşmaktadır.

3.1.2 Mekan ve kullanıcı verisi hazırlama

Çalışmada kullanılan mekan verisi ve bu mekanları ziyaret etmiş kullanıcılardan oluşan kullanıcı verisi geliştirilen ayrıştırıcı ile oluşturulmuştur. Geliştirilen ayrıştırıcı *Jericho HTML Parser* [39] kütüphanesini kullanmaktadır. Web sitesindeki restoran ve kullanıcılara ait ilgili içerik bilgileri ayrıştırıcı ile veritabanı için uygun formata çevrilerek oluşturulan restoran veritabanı ve kullanıcı veritabanına aktarılmıştır. Web sitesindeki örnek bir restoranın içerik bilgileri Şekil **3.2**'deki gibi gösterilmiştir.

Restoran veritabanı, İstanbul'un çeşitli semtlerinde bulunan restoran tarzı mekanların özellik bilgilerinden elde edilmiştir. Kullanıcı veritabanı ise bu mekanları ziyaret etmiş ve mekanlar hakkında yorum yazmış kullanıcıların özellik bilgilerinden oluşmaktadır. Her iki veritabanını oluşturmak için öncelikle bir ayrıştırıcı (parser) geliştirilmiş ardından Web ortamında mekanların ve bu mekanlara giden kullanıcıların bulunduğu Web sitesindeki bilgiler ayrıştırıcı ile toplanarak veritabanına kayıt edilmiştir.



Şekil 3.2 : Ayrıştırıcının kullanıldığı Web sayfası.

Restoranların özellik bilgilerinden oluşan restoran veritabanı; restoran ismi, fiyat aralığı, giyim tarzı, rezervasyon, kategori, bulunduğu semt ve konum bilgilerinden oluşmaktadır. Kullanıcıların özellik bilgilerinden oluşan kullanıcı veritabanı; kullanıcı ismi, kullanıcı eğitim durumu, meslek ve kullanıcının gittiği mekan bilgilerinden oluşmaktadır. Ayrıştırma işlemi sonucunda 725 restoran ve bu mekanlara giden 551 kullanıcının özellik bilgilerinden oluşan restoran ve kullanıcı veritabanları oluşturulmuştur. Kullanıcı sayısının restoran sayısından az olmasının nedeni bazı kullanıcıların birden fazla restoranı ziyaret etmesi ve bazı restoranlarında sadece bir kullanıcı tarafından ziyaret edilmesidir.

3.2 Rasgele Yürüyüş Algoritması Kullanılarak Topluluk Belirleme Yöntemi

Kullanıcıların konuşma kayıtlarından oluşan konuşma verisi yönsüz ağırlıklı bir çizge $G = (V,E)$ ile temsil edilmiştir. Çizgenin düğümleri (V) mobil telefon kullanıcılarını, ayrıtları ise (E) konuşma sürelerini temsil etmektedir. w kişisi u kişisini ya da u kişisi w kişisini aramışsa w ile u düğümleri arasında ağırlığı w ile u arasındaki toplam arama süresi olan yönsüz bir ayrıt oluşmaktadır.

Tez çalışması kapsamında daha çok konuşan kullanıcılardan oluşan topluluk yapılarının belirlenmesi için gerçeklenen yöntem Rasgele Yürüyüş Yöntemi Kullanılarak Topluluk Belirleme Yöntemidir [2]. Bu yöntem düğümler arasında rasgele yürüyüş tabanlı benzerlik ölçümünü kullanmaktadır. Benzerlik ölçümü olarak rasgele yürüyüş yönteminin kullanılmasının birçok avantajı bulunmaktadır. Bu yöntem sayesinde çizge içindeki topluluklar verimli bir şekilde bulunmakta ve algoritmanın karmaşıklığı, n düğüm sayısı olmak üzere, $O(n^2 log\ n)$ olarak belirlenmekte, bu nedenle düğüm sayısı fazla olan ağlarda verimli sonuçlar elde edilmektedir. Algoritmanın gerçeklenmesi sonucunda amaçlanan, ağırlık değerleri büyük düğümlerin birleşiminden oluşan gruplara topluluk adı verilerek bu düğüm gruplarının bulunmasıdır.

Topluluk Belirleme Algoritması ilk önce bütün düğümleri ve ağırlıklarını dosyadan okuyarak çizge yapısını oluşturmaktadır.

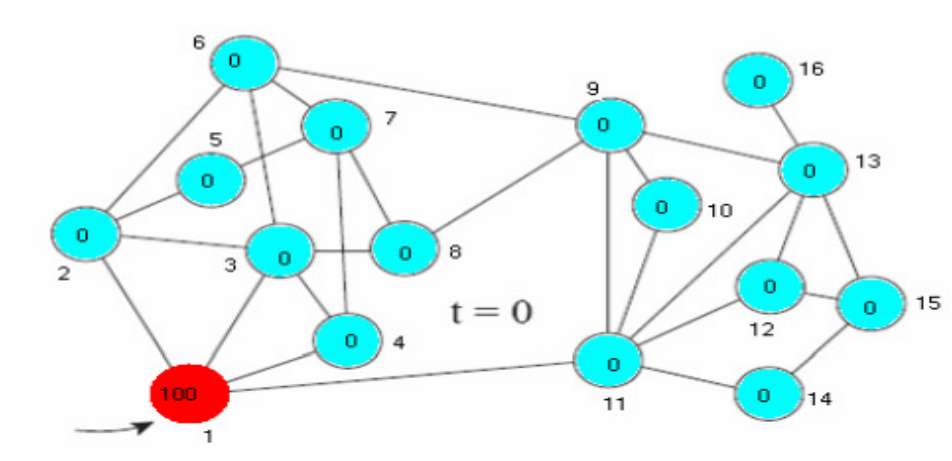
$$P_{t+1}(j) = \sum_{i=1}^{j} (P_t(i) \frac{A_{ij}}{d(i)}) \quad \textbf{(3.1)}$$

Ardından her düğüm birer topluluk kabul edilerek olasılık değerleri Eşitlik **(3.1)**'deki gibi hesaplanmaktadır. Olasılık değerleri hesaplanırken algoritma sırasıyla her bir düğümden başlayarak bütün düğümleri gezmekte ve bütün düğümlere gitme olasılığını her bir düğüm için hesaplamaktadır. Algoritma bu sayede çizge yapısının topolojisi hakkında fikir edinmektedir. *j* olasılık değeri hesaplanacak düğümü; $P_{t+1}(j)$, *j.* düğümünün olasılık değerini; *Aij, i* ve *j* düğümleri arasındaki ağırlık değerini; *d(i), i.* düğümünün derecesini; $P_t(i)$, *i.* düğümünün olasılık değerini göstermektedir. *t* değeri algoritmanın giriş parametresi olarak aldığı rasgele yürüyüşün iterasyon sayısıdır. *j.* düğümünün olasılık değeri, *i.* düğümden *j.* düğüme geçiş olasılığını göstermekte ve Eşitlik **(3.1)**'deki gibi hesaplanmaktadır.

Ardından her düğüm için birer olasılık vektörü oluşturulmakta ve bu olasılık vektörleri kullanılarak düğümler arasındaki uzaklık değerleri Eşitlik **(3.2)**'deki gibi bulunmaktadır.

$$r_{ij} = \sqrt{\sum_{k=1}^{n} \frac{(P_{ik}(t) - P_{jk}(t))^2}{d(k)}} \quad \textbf{(3.2)}$$

r_{ij}, *i* ve *j* düğümleri arasındaki uzaklık değerini; P^t_{ik} ve P^t_{jk}, *i.* ve *j.* düğümlerin, *t* iterasyon adımındaki olasılık vektörlerini göstermektedir.



Şekil 3.3 : t=0 için G çizge yapısı [2].

Şekil 3.3'teki örnek *G* çizgesi üzerinden algoritmayı anlatmak gerekirse, örnek çizge 16 düğümden oluşmaktadır ve düğümler arasındaki ayrıtların ağırlığı notasyonu basitleştirmek için 1 olarak kabul edilmiştir. *G* çizgesi, Çizelge **3.1**'deki gibi *A* komşuluk matrisi ile temsil edilmektedir. Örnek olarak 1. düğüme olasılık başlangıç değeri olarak 100 atanmıştır.

Çizelge 3.1 : Komşuluk Matrisi A.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0
2	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0
3	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0
4	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
5	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0
6	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0
7	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0
8	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0
9	0	0	0	0	0	1	0	1	1	1	1	0	1	0	0	0
10	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
11	1	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0
12	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0
13	0	0	0	0	0	0	0	0	1	0	1	1	1	0	1	1
14	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0
15	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0
16	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1

Eğer *i* ve *j* düğümleri arasında ayrıt bulunuyorsa, Çizelge **3.1**'deki komşuluk matrisindeki *Aij* değerleri, *i* ve *j* düğümleri arasındaki ayrıtın ağırlık değeri olmaktadır. Notasyonu basitleştirmek için Şekil **3.3**'te verilen örnek çizgede ayrıtların ağırlığı 1 olarak kabul edilmiş ve buna göre *A* komşuluk matrisi oluşturulmuştur. Eğer *i* ve *j* düğümleri arasında ayrıt bulunuyorsa *Aij* değeri 1 diğer durumda 0'dır. *d(i)*, *i* düğümünün derecesini göstermekte ve *i* düğümünün komşularının sayısını belirtmektedir. Düğümlere ait derece değerleri Çizelge **3.2**'deki derece tablosunda gösterilmektedir.

Çizelge 3.2 : Derece Tablosu d.

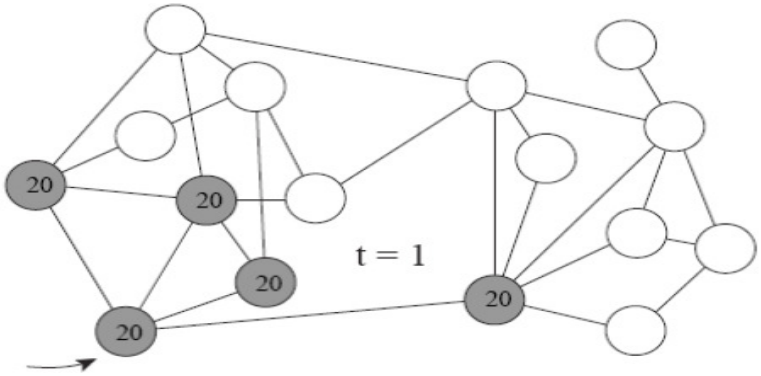
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
5	5	6	5	3	5	5	4	6	3	7	4	6	3	4	2

Rasgele yürüyüşün her bir iterasyon adımında bulunulan düğümden komşular arasından rasgele seçilmiş bir başka düğüme geçilmektedir. Her bir adımda *i* düğümünden *j* düğümüne geçiş olasılığı olan *Pij* değerleri Eşitlik **(3.1)**'deki gibi hesaplanarak rasgele yürüyüşün geçiş (olasılık) matrisi *P* bulunmaktadır.

Çizelge 3.3 : Geçiş Matrisi P t = 0 için.

100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Şekil **3.3**'teki örnek *G* çizgesinde görüldüğü gibi 1. düğüme başlangıç olasılık değeri olarak 100 atanmış ve $t = 0$ için *P* geçiş matrisi Çizelge **3.3**'teki gibi oluşturulmuştur.



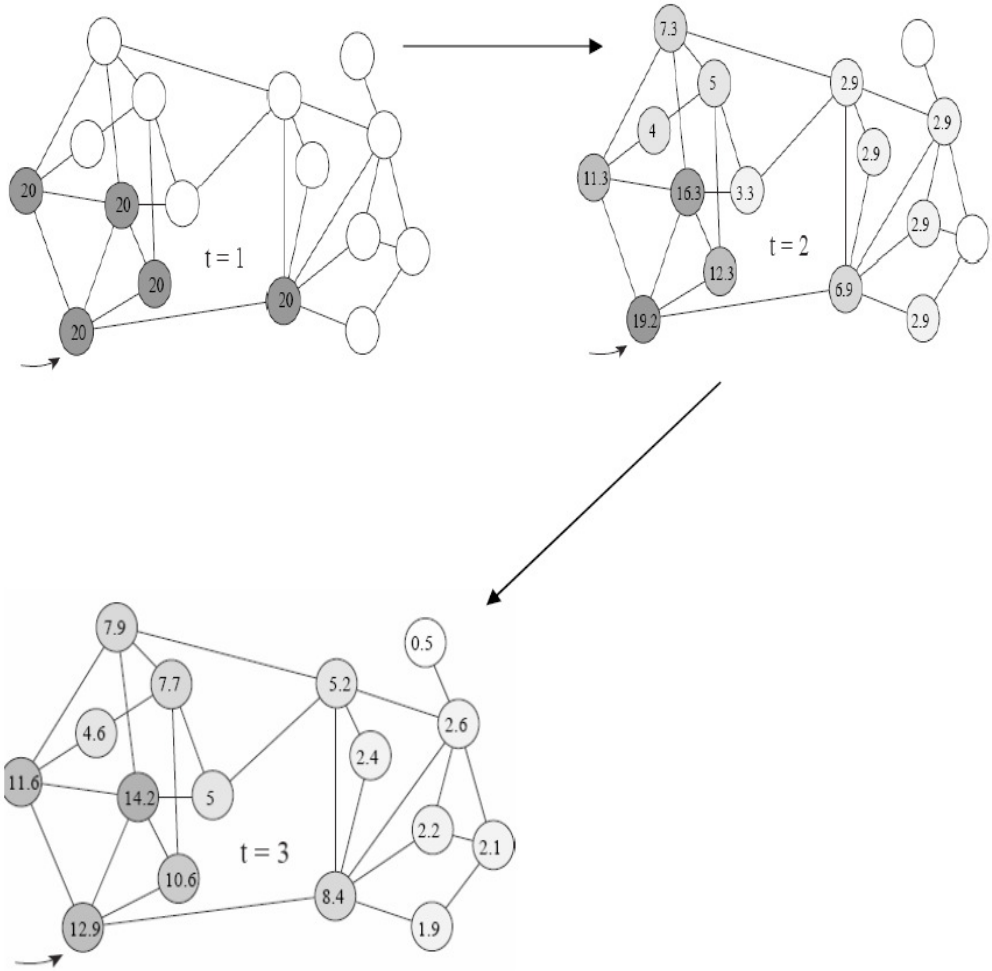
Şekil 3.4 : Geçiş Çizgesi t=1 için [2].

Ardından $t = 1$ iterasyon adımına geçilmekte ve bu adım için *i* düğümünden *j* düğümüne geçiş olasılığı olan *Pij* değerleri Eşitlik **(3.1)**'deki gibi hesaplanarak rasgele yürüyüşün geçiş matrisi *P*, Çizelge **3.4**'deki gibi güncellenmektedir. $t = 1$ iterasyon adımı için bulunan olasılık değerleri Şekil **3.4**'teki gibi çizge üzerinde de gösterilmiştir.

Çizelge 3.4 : Geçiş Matrisi P t = 1 için.

20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

t değerinin, algoritmanın giriş parametresi olarak aldığı rasgele yürüyüşün iterasyon sayısı olduğu daha önce belirtilmişti, *t*=1, *t*=2 ve *t*=3 için olasılık değerleri Şekil **3.5**'teki gibi bulunmuş ve her aşamada *P* geçiş matrisi güncellenmiştir. Üç iterasyon sonunda algoritma bütün düğümler için olasılık değerlerini Eşitlik **(3.1)**'deki formüle göre bulmakta ve bu değerlere göre düğümler arasındaki uzaklık değerleri Eşitlik **(3.2)**'deki formüle göre belirlenmektedir. Ardından hiyerarşik demetleme algoritması uygulanarak topluluklar belirlenmektedir.



Şekil 3.5 : Geçiş Çizgesi t=3 için [2].

Algoritmada topluluk yapılarının bulunması için hiyerarşik demetleme algoritması gerçeklenmiştir. Düğümler arasındaki uzaklığa göre demetleme işlemi gerçekleştirilmektedir. Algoritma başlangıç olarak bütün düğümleri bir topluluk olarak kabul etmekte ardından demetleme işlemi sonucunda topluluk sayısını bire indirmektedir. Algoritmada ilk olarak bütün komşu düğümler arasındaki uzaklıklar Eşitlik **(3.2)**'deki gibi bulunmakta ardından aşağıdaki iterasyon adımları gerçeklenmektedir. Sınıf diyagramı ve sınıflara ilişkin açıklamalar **EK A.1**'deki gibi verilmiştir.

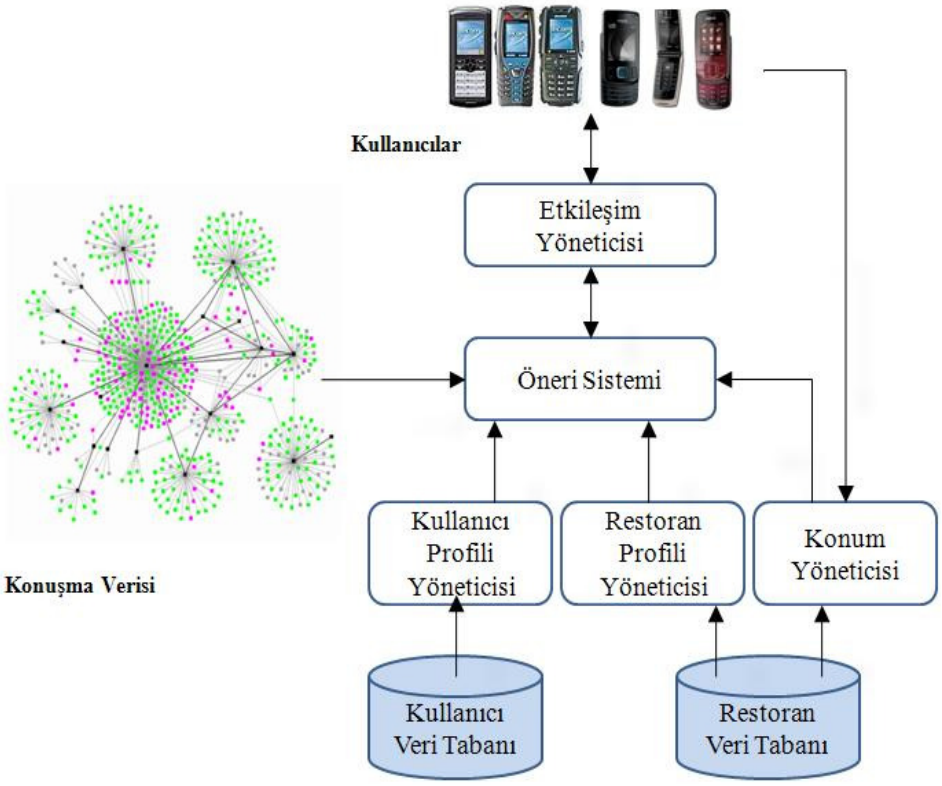
1. Uzaklık ve olasılık matrisini hesapla
2. Her düğüm bir topluluk
3. Tekrarla (Kuyruktaki elemanlar bitene kadar)
4. En yakın iki topluluğu birleştir
5. Uzaklık ve olasılık matrisini yeniden hesapla ve kuyruğu güncelle
6. Sonlanma: Tek bir demet kalana kadar

Algoritma bulunan düğümleri topluluklara gruplamak için düğümler arasındaki uzaklığı kullanmaktadır. Algoritmanın ilk adımında topluluklar arasındaki uzaklık değerleri hesaplanmaktadır. Ardından oluşturulan topluluk çiftleri için birer eleman oluşturulmakta ve bu elemanlar kuyruğa atılmaktadır. Kuyrukta bu elemanlar uzaklık değerlerine göre sıralanmıştır. Bu kuyruk yapısı içinde hiyerarşik demetleme yöntemi kullanılarak topluluk çiftleri birleştirilmekte bu birleştirme işlemi tek bir topluluk kalıncaya kadar devam etmektedir. Algoritma sadece olasılık değerlerinin hesabı sırasındaki iterasyon sayısını parametre olarak almaktadır.

3.3 Öneri Altyapı Mimarisi

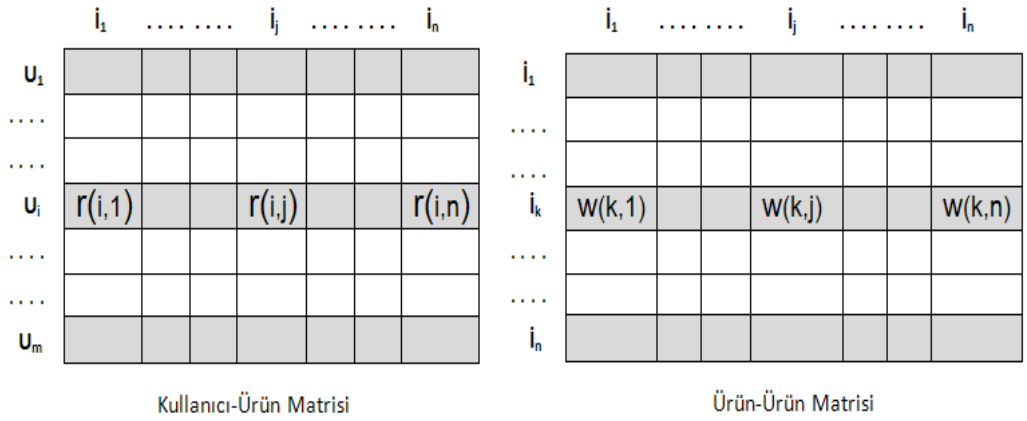
Genel sistem mimarisi Şekil **3.6**'daki gibi gösterilmiştir. Tasarlanan altyapı mimarisi etkileşim yöneticisi, öneri sistemi, konum yöneticisi, restoran profili yöneticisi ve kullanıcı profili yöneticisi olmak üzere beş ana gruba ayrılmıştır. Etkileşim yöneticisi mobil telefon kullanıcıları ve öneri sistemi arasındaki arayüzü oluşturmaktadır. Kullanıcılardan öneri isteklerini almaktan ve kullanıcılara öneri sistemi tarafından üretilen öneri mesajlarını SMS, mobil Web sayfası gibi mobil iletişim kanalları aracılığı ile sunmaktan sorumludur. Konum yöneticisi ise Küresel Yer Belirleme Sistemi (GPS) ile kullanıcıya ait coğrafi konum, zaman ve cihaz bilgisi gibi ortam bilgilerini toplamaktan ve kullanıcının konumunun yakınındaki mekanları belirlemekten sorumludur. Kullanıcı profil yöneticisi, kullanıcı veritabanı ve restoran veritabanını kullanarak kullanıcıların ziyaret örüntülerine göre kullanıcı profillerini oluşturmaktadır. Sonuç olarak oluşturulan kullanıcı profilleri kullanıcıların mekan tercihlerini tanımlamaktadır. Restoran profil yöneticisi, kullanıcıların tercih ettiği restoranlara göre restoran profillerini tanımlamaktadır. Son olarak öneri sistemi etkileşim yöneticisinden öneri isteklerini almakta ve içerik filtreleme, işbirlikçi filtreleme ve sosyal ağ analizi yöntemlerini kullanarak restoran önerilerini üretmektedir. Belirli bir kullanıcı için öneriler, sosyal ağda aynı toplulukta bulunan kullanıcılar arasındaki benzerlikler ve restoranlar arasındaki benzerlikler hesaplanarak üretilmektedir.

Bu çalışmada iki veri kaynağı kullanılmaktadır. Birinci veri kaynağı yönsüz ağırlıklı çizge olarak temsil edilen konuşma verisinden oluşmaktadır. Topluluk belirleme işlemi kullanıcı profil yöneticisi tarafından gerçeklenmektedir. İkinci veri kaynağı kullanıcıların tercih bilgisinden ve restoranların içerik bilgisinden oluşmaktadır. Restoranların özellikleri ve kullanıcıların restoran ziyaret bilgileri, çeşitli mekan içerik bilgilerinden oluşan bir Web sitesinden elde edilmiştir.



Şekil 3.6 : Öneri altyapı mimarisi.

Kullanıcıların restoran ziyaret bilgileri ve restoranların özellik bilgileri kullanılarak Kullanıcı-Ürün matrisi ve Ürün-Ürün matrisi oluşturulmuştur. *m* satır ve *n* sütundan oluşan Kullanıcı-Ürün matrisinde *m* kullanıcı sayısını *n* ürün sayısını göstermektedir, her bir ürün restoran veritabanındaki bir restorana karşı gelmektedir. Kullanıcı-Ürün matrisinin her bir *r(i,j)* değeri kullanıcı u_i'nin restoran i_j'yi kaç defa ziyaret ettiğini belirtmekte ve bu değerin büyük olması kullanıcının ilgili restoranı çok fazla sayıda tercih ettiğini göstermektedir. *r(i,j)* değerinin 0 olması ise kullanıcı u_i'nin restoran i_j'yi ziyaret etmediğini veya ilgi duymadığını göstermektedir. Ürün-Ürün matrisinin değerleri ise restoran çiftleri arasındaki benzerliği göstermektedir. Kullanıcı-Ürün ve Ürün-Ürün matrisleri Şekil **3.7**'de gösterilmiştir.



Şekil 3.7 : Kullanıcı-Ürün ve Ürün-Ürün matrisi.

Restoran profil yöneticisinde, fi_j değeri *Ri* restoranının bir özelliğini temsil etmekte olup her bir restoran $\boldsymbol{Ri} = (fi_1, fi_2, \ldots, fi_k)$ özellik vektörü şeklinde temsil edilmekte ve restoran çiftleri arasındaki benzerlik hesaplanmaktadır. Örnek olarak türü gece klübü, ortalama maliyeti 30 TL, ödeme tipi kredi kartı, giyim tarzının günlük, rezervasyon şartının gerekmediği ve bulunduğu yerin Taksim olduğu örnek bir restoranın özellik vektörü altta görüldüğü gibi temsil edilmektedir.

$$\boldsymbol{Ri}=(\{gece\ klübü\},30,\{kredi\ kartı\},\{günlük\},\{reservasyon\ gerekmiyor\},\{Taksim\})$$

i_k ve i_j restoranları arasındaki benzerlik ifadesi olan $w(k,j)$ değeri Eşitlik **(3.3)**'deki gibi bulunmaktadır.

$$w(k,j) = weight1 \times sim_{cos}(i_k, i_j) + weight2 \times |i_k(u) \cap i_j(u)| \qquad \textbf{(3.3)}$$

$i_k(u)$ ve $i_j(u)$, i_k ve i_j restoranlarını tercih eden kullanıcıların kümesini temsil etmekte, $|i_k(u) \cap i_j(u)|$ değeri de hem i_k hem de i_j restoranını tercih eden kullanıcıları belirtmektedir. *weight1,* her iki restoran arasındaki kosinüs benzerliğinin ağırlık değerini; *weight2*, her iki restorana giden kullanıcı sayısının ağırlık değerini göstermektedir. *weight1*+ *weight2* = 1 olacak şekilde deneysel çalışmalarda farklı ağırlık değerleri belirlenmiş ve öneri modelinin sonuçları incelenmiştir.

Aynı topluluk içinde bulunan u_x ve u_y kullanıcıları arasındaki benzerlik değeri Eşitlik **(3.4)**'deki gibi bulunmaktadır.

$$sim(Ux,Uy) = \sum_{i=1}^{n} \sum_{j=1}^{n} r(x,i)r(y,j)w(i,j) \quad \textbf{(3.4)}$$

Kullanıcı profil yöneticisi tarafından her bir u_x kullanıcısı için kullanıcı profili, $u_x profile = (fu_{x1}, fu_{x2}, \ldots, fu_{xk})$, kullanıcısının tercih ettiği restoranların özellik vektörlerinden oluşturulmaktadır. Özellik vektöründeki fu_{xi} değeri, u_x kullanıcısının tercih ettiği restoranların özellik vektörlerindeki *i.* özelliklerinin ortalamasıdır. Restoranların özellik vektöründeki sayısal olmayan kategorik özellikler için en fazla bulunan özellik değeri ortalama değer olarak alınmıştır.

Kullanıcı u_x ve u_y sosyal ağda aynı topluluk içinde bulunan kullanıcılardır. *Ri* restoranı kullanıcı u_y tarafından tercih edilip kullanıcı u_x tarafından tercih edilmeyen bir restorandır. Öneri sisteminde kullanıcı u_x'e önerilecek *Ri* restoranı için öneri puanı Eşitlik **(3.5)**'teki gibi hesaplanmaktadır.

$$recScore\ (Ux,Ri)\ = \frac{1}{|\,S(Ri)\,|} \sum_{Uy \in S(Ri)} sim\,(Ux,Uy)\,sim\,(Uxprofile\ ,Ri) \quad \textbf{(3.5)}$$

$sim(u_x profile, Ri)$ kullanıcı u_x'in profili ile *Ri* restoranının özellik vektörü arasındaki benzerliği ifade etmektedir. *S(Ri)* kümesi ise u_x ile aynı toplulukta olup *Ri* restoranını tercih eden kullanıcıların kümesini ifade etmektedir.

İlgili kullanıcıdan öneri isteği geldiği zaman aynı toplulukta bulunduğu diğer kullanıcılar ile ilgili kullanıcının benzerlik değerleri hesaplanmakta ardından diğer kullanıcıların gittiği ilgili kullanıcının gitmediği mekanlar için öneri skoru hesaplanmaktadır. Kullanıcının konumuna göre mekanlar filtrelenmekte ardından en yüksek öneri skoruna sahip 5 mekan ilgili kullanıcıya önerilen restoranlar olarak sunulmaktadır.

3.4 Mobil Telefon Simülatörü

Geliştirilen altyapı mimarisinin kullanımını modellemek için mobil telefon simülatörü tasarlanmış ve kullanıcılara mobil telefon arayüzü aracılığı ile mekan önerileri sunulmuştur. Mobil telefon simülatörü, mobil cihazlara Java uygulaması geliştirmek için kullanılan standart olan J2ME ortamında geliştirilmiştir.

Mobil telefon simülatörü ile öneri altyapısı arasındaki iletişim; istemci sunucu mimarisinde, Tomcat Sunucusu üzerinde Basit Nesne Erişim Protokolü (SOAP) Web Servisleri kullanılarak gerçeklenmiştir. Web Servis teknolojisi uzaktaki bir sistemde var olan hizmete ağ üzerinden erişimi sağlamaktadır. Bu sayede istemci ve sunucu birbirinden bağımsız geliştirilebilmektedir. Basit Nesne Erişim Protokolü ile Web Servisin sağlamış olduğu fonksiyon uygun parametrelerle uzaktan çağrılabilmekte, istemciye sunucu tarafından elde edilmiş sonuçlar gönderilmektedir.

Basit Nesne Erişim Protokolü (SOAP), Genişletilebilir İşaretleme Dili (XML) tabanlı uygulamalarda Hiper Metin Transfer Protokolü (HTTP) üzerinden bilgi alışverişine izin veren Web Servis erişim protokolüdür. Web Servislere, diğer uygulamalardan XML ile erişilebilir bu sayede uygulama ve programlama dilinden bağımsız olarak geliştirmeler yapılabilmektedir. Bunun için Web Servis Tanımlama Dili (WSDL), XML tabanlı Web Servisleri tanımlamak için kullanılmaktadır. WSDL, Web Servisin mesaj formatlarını ve ağ adresi gibi bilgilerini tanımlamaktadır [40] [41].

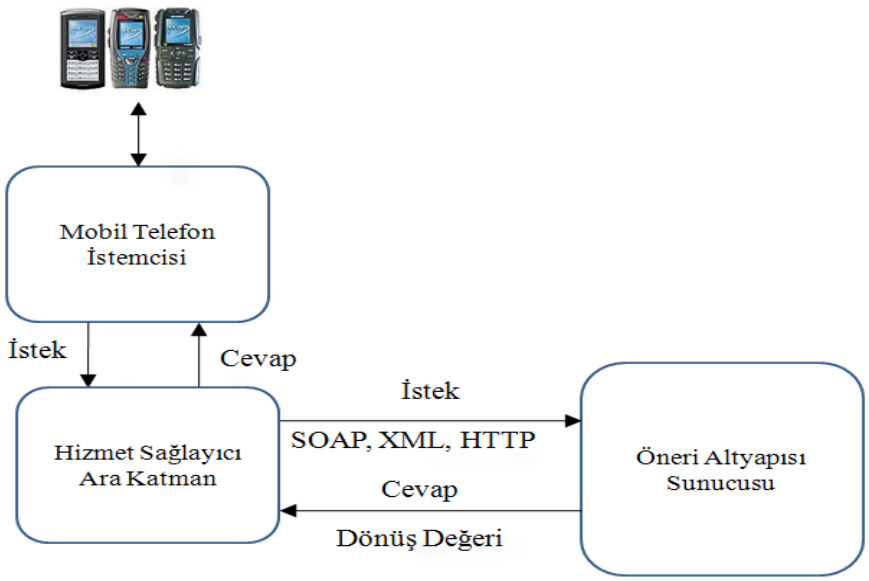
Tasarım geliştirme sürecinde ihtiyaç duyulan yazılımlar ve araçlar alttaki gibi verilmiştir.

- Sun JDK 1.5.02 (*Sun Java uygulama geliştirme kiti*)
- Sun Java Wireless Toolkit 2.5.2 (*Mobil uygulama geliştirme kiti*)
- Eclipse 3.2.1 (*Java ile uygulama geliştirme ortamı*)
- Tomcat 6.0 Server (*Uygulama sunucusu*)
- MySQL 5.1 (*Açık kaynaklı veritabanı*)

Web Servisin istemcisi olan mobil telefon simülatörü, sunucu olan öneri altyapısındaki servisi çağırma aşamasındaki temel adımları alttaki gibi verilmiştir.

1. Web Servis istemcisi olan mobil telefon simülatörü, XML tabanlı web servisi bulur.
2. İstemci bir SOAP mesajı hazırlar. SOAP mesajı bir XML belgesidir.
3. İstemci SOAP mesajını uygulama sunucusundaki çalışan SOAP istek dinleyicisine gönderir.
4. Sunucu gelen SOAP mesajını ayrıştırır ve gerekli parametreleri göndererek istenen nesnenin istenen yordamını çağırır.
5. Çağrılan nesnedeki yordam çalışır ve sonuçları uygulama sunucusuna gönderir. Uygulama sunucusu gelen sonuçları SOAP mesajı formatında biçimlendirerek istemciye gönderir.
6. Web Servis istemcisi olan mobil telefon simülatörü, SOAP mesajının içindeki bilgileri alarak ekran arayüzünde kullanıcıya göstermektedir.

Hem sunucu hem de istemci Java tabanlı uygulamalardır. Şekil **3.8**'de görüldüğü gibi istemci tarafını, mobil istemci olan mobil telefon simülatörü; sunucu tarafını ise Web Servis teknolojisi kullanılarak gerçeklenmiş öneri altyapısı sunucusu oluşturmaktadır.



Şekil 3.8 : İstemci-Sunucu mimarisi [42].

İstemci ile sunucu arasındaki iletişimi sağlamak için üç tane Web Servis geliştirilmiştir. Web Servislerin gerçeklendiği Web Sunucusu olarak Tomcat 6.0 kullanılmış; arama yapmak, önerilen mekanları görmek ve oylama işlemi yapmak için mobil istemcinin gönderdiği öneri isteklerine cevaplar üretilmiştir. Örnek olarak kullanıcı ana giriş ekranı ve öneri görme ekranı Şekil **3.9**'da, önerilen mekanları oylama ekranı da Şekil **3.10**'da gösterilmiştir.



Şekil 3.9 : Kullanıcı ana giriş ekranı ve öneri görme ekranı.



Şekil 3.10 : Önerilen mekanları oylama ekranı.

Öneri altyapısı tarafından üretilen restoran önerileri veritabanına her kullanıcı için kayıt edilmektedir ardından kullanıcıya mobil telefon aracılığı ile yapılan öneriler sunulmaktadır. Kullanıcı mobil telefon aracılığı ile öneri isteğinde bulunmakta ve kullanıcının profiline uygun olarak üretilen restoran önerileri bulunarak mobil telefon arayüzü aracılığı ile mekan önerileri sunulmaktadır. Web Servis teknolojisi kullanılarak uzaktan çağrılabilen sunucu, sunucunun yordamını çağırarak çalışan mobil istemcilerdir. Java diliyle yazılmış olması nedeniyle platform bağımsız olan proje Windows işletim sisteminde geliştirilmiştir.

4. DENEYSEL SONUÇLAR

Bu bölümde, Bölüm 3.3'te tanıtılan Sosyal Ağ Verisi Kullanan Mobil Öneri Sisteminin oluşturulması aşamaları ve veri kümeleri üzerindeki önermelere ait sonuçlar sunulmaktadır. Sonuçların genel geçerliliği, ülkemizde hizmet veren bir mobil servis operatöründen elde edilen konuşma kayıtlarına ait veri kümesi ve bir Web sayfasından elde edilen restoran içerik bilgilerini içeren veri kümesi üzerinde sınanmıştır. Aşağıdaki sonuç başlıklarında her iki veri kümesi için de sonuçlar verilmektedir.

4.1 Veri Hazırlama Sonuçları

Bölüm 3.1'de aktarılan veri hazırlama işlemleri olan sosyal ağ verisi hazırlama, mekan kullanım verisi hazırlama ve kullanıcı eşleştirme sonuçlarından elde edilen ölçümler bu bölümde verilmektedir.

Tez çalışmasında tasarlanan modelin başarımını değerlendirmek için yapılan deneysel çalışmada, ülkemizde hizmet veren mobil servis sağlayıcısından elde edilen konuşma verisi ve bir Web sitesinden elde edilen restoran ve kullanıcı verileri olmak üzere iki tane veri kümesi kullanılmıştır. Restoran veri kümesi İstanbul'da bulunan 725 adet restorandan oluşmaktadır. Restoran veri kümesinin 70%'i sunulan modeli eğitmek için öğrenme kümesi olarak ayrılmış, veri kümesinin geri kalan kısmı da test kümesi olarak belirlenmiştir. Konuşma verisindeki 551 kullanıcı için konuşma çizgesi oluşturulmuş ve bu çizge üzerinde rasgele yürüyüş algoritması uygulanarak birbirleriyle sık konuşan kullanıcılar demetlenmiş ve on tane topluluk yapısı belirlenmiştir. Öğrenme kümesindeki restoranlar Kullanıcı-Ürün ve Ürün-Ürün Matrislerinin değerlerinin bulunmasında ve kullanıcılara öneri yapılmasında kullanılmıştır. Yapılan önerilerin doğruluğu test kümesi üzerinde sınanmıştır.

Günümüzde mobil telefonlar aracılığı ile kişilerin ne zaman nereyi ziyaret ettiği bilgisi bulunabilmektedir. Fakat bu tarz bilgilerin paylaşılması kişilere ait özel ve gizli bilgi olduğu için paylaşılması mobil servis operatörleri için yasaktır. Özel hayatın korunması ve kişisel gizlilik ilkesi gereği gibi nedenlerle bu çalışmada kişilerin gerçek restoran ziyaret bilgilerini içeren bir veri kümesi elde edilememiştir.

Mobil servis operatöründen elde edilen konuşma verisindeki kullanıcılar, bir Web sitesinden elde edilen İstanbul içindeki restoranlara giden kullanıcılarla rassal olarak eşleştirilerek tasarlanan öneri altyapı mimarisinin tutarlı bir şekilde çalışması deneysel çalışmalarla gösterilmiştir. Birbirleriyle sık konuşan kişilerin benzer mekanlara gitme eğiliminde olduğuna dair ön sonuçlar Bölüm 4.2'de incelenecektir.

Konuşma çizgesindeki 551 kullanıcı için rasgele yürüyüş ile topluluk belirleme algoritması uygulanmış ve birbirleriyle sık konuşan kullanıcılardan oluşan on topluluk belirlenmiştir. Belirlenen on topluluktan dokuz tanesi 55 kullanıcı, bir tanesi 56 kullanıcı içermektedir.

4.2 Öneri Altyapısı Sonuçları

Bölüm 3.1'de aktarılan veri hazırlama işlemleri olan sosyal ağ verisi hazırlama, mekan kullanım verisi hazırlama ve kullanıcı eşleştirme sonuçlarından elde edilen ölçümler bu bölümde verilmektedir.

Deneysel çalışmaların ilk kısmı için 10 farklı veri kümesi oluşturulmuş ardından tasarlanan öneri altyapısı sınanmıştır. 10 farklı veri kümesinin (*DSi, i* = 1,…,10, *Çizelge* ***4.1***, *Çizelge* ***4.2***, *Çizelge* ***4.3,*** *Çizelge* ***4.4***) her biri eşit sayıda kullanıcıyı içermektedir. Her bir veri kümesi, toplulukların her birinden rassal olarak alınan 30 kullanıcıdan oluşmaktadır. Veri kümesinin toplamı 300 kullanıcıdan oluşmaktadır.

Her bir veri kümesi için alttaki adımlar tasarlanan öneri altyapısını sınamak ve başarımını ölçmek için gerçeklenmiştir.

İlk adım olarak veri kümesindeki bütün kullanıcılar için, her bir kullanıcının ziyaret ettiği restoranlar öğrenme ve test kümesi olmak üzere iki parçaya ayrılmıştır. Öğrenme ve test kümelerinin oluşturulması aşamasında her bir kullanıcı için; kullanıcının ziyaret ettiği restoranların %70'i rassal olarak seçilerek öğrenme kümesi, kalan restoranlarda test kümesi olarak belirlenmiştir.

İkinci adım olarak her bir kullanıcı için; kullanıcı u_x'in profili olan *uxprofile*, ilgili kullanıcının öğrenme kümesindeki restoranlardan oluşturulmaktadır.

Üçüncü adım olarak önerilerin yapılacağı kullanıcı, aktif kullanıcı olarak isimlendirilmekte; aktif kullanıcı ile aynı toplulukta bulunan ve Eşitlik (**3.4**)'deki benzerlik değerine göre aktif kullanıcıya en benzer beş kullanıcı belirlenmektedir.

Dördüncü adımda belirlenen beş kullanıcının ziyaret ettiği, aktif kullanıcının ziyaret etmediği restoranlar için Eşitlik (**3.5**)'deki gibi öneri puanı hesaplanmakta ve yüksek puan almış olan restoranlar aktif kullanıcıya önerilmek üzere belirlenmektedir. Her bir aktif kullanıcı için yapılan öneri sayısı, kullanıcının test kümesindeki restoran sayısının %33'ü kadardır. Eğer önerilen bir restoran aktif kullanıcının test kümesinde bulunursa; bu restoran önerisi aktif kullanıcı için doğru bir öneri olarak belirlenmektedir.

4.2.1 Kesinlik, anma ve f-ölçütü sonuçları

Belirli bir veri kümesi için Kesinlik değerinin hesaplanmasında, veri kümesinde bulunan bütün kullanıcılar için belirlenen doğru önerilerin sayısı, toplam yapılan restoran öneri sayısına oranlanmaktadır. Kesinlik değerinin hesaplanması Eşitlik (**4.1**)'deki gibi verilmiştir. $|\textit{Önerilen Restoranlar} \cap \textit{Test Kümesi Restoranları}|$ ilgili veri kümesinde kullanıcılar için yapılan doğru önerilerin sayısını, $|\textit{Önerilen Restoranlar}|$ ilgili veri kümesi için yapılan bütün önerilerin sayısını göstermekte, oranları da Kesinlik değerini vermektedir.

$$\text{Kesinlik} = \frac{|\,\text{Önerilen Restoranlar} \cap \text{Test Kümesi Rest.}\,|}{|\,\text{Önerilen Restoranlar}\,|} \qquad \textbf{(4.1)}$$

Diğer yandan belirli bir veri kümesi için Anma değerinin hesaplanmasında, veri kümesinde bulunan bütün kullanıcılar için belirlenen doğru önerilerin sayısı, veri kümesindeki bütün kullanıcılara ait test kümelerindeki toplam restoran sayısına oranlanmaktadır. Anma değerinin hesaplanması Eşitlik (**4.2**)'deki gibi verilmiştir.

$|\textit{Test Kümesi Restoranları}|$ ilgili veri kümesindeki bütün kullanıcılara ait test kümelerindeki toplam restoran sayısını göstermekte, doğru önerilerin sayısı ile oranları da Anma değerini vermektedir.

$$\text{Anma} = \frac{|\text{Önerilen Restoranlar} \cap \text{Test Kümesi Rest.}|}{|\text{Test Kümesi Rest.}|} \quad \textbf{(4.2)}$$

Veri kümelerindeki bütün kullanıcılar için öneri işlemi yapıldıktan sonra her bir veri kümesi için Kesinlik, Anma ve F-Ölçütü metrikleri hesaplanmaktadır. F-Ölçütünün bulunmasında Eşitlik (**4.3)**'de olduğu gibi Anma ve Kesinlik değerlerinin harmonik ortalaması alınmıştır.

$$\text{F - Ölçütü} = \frac{2 \text{ x Kesinlik x Anma}}{\text{Kesinlik + Anma}} \quad \textbf{(4.3)}$$

Öneri altyapısının sınanması için oluşturulan, 10 veri kümesi için hesaplanan Kesinlik, Anma ve F-Ölçütü değerleri; Eşitlik (**3.3)**'teki farklı ağırlık değerleri için Çizelge **4.1**, Çizelge **4.2**, Çizelge **4.3**'deki gibi verilmiştir.

Çizelge **4.1**'de verilen sonuçta, restoranlar arasındaki benzerlik ifadesi olan Eşitlik (**3.3)**'teki *weight1* ve *weight2* değerleri 0.5 olarak alınmıştır.

Çizelge 4.1 : Tasarlanan Öneri Modeli Sonuçları (weight1 = 0.5, weight2 = 0.5).

Tasarlanan Altyapı Modeli	Kesinlik	Anma	F-Ölçütü
DS1	0.6488	0.2092	0.3164
DS2	0.5858	0.1926	0.2901
DS3	0.6697	0.2188	0.3299
DS4	0.6796	0.2250	0.3381
DS5	0.6059	0.2013	0.3022
DS6	0.6082	0.1996	0.3006
DS7	0.6165	0.2041	0.3067
DS8	0.6199	0.2038	0.3068
DS9	0.6666	0.2201	0.3309
DS10	0.6682	0.2192	0.3302
Ortalama	0.6369	0.2094	0.3152

Çizelge **4.2**'de verilen sonuçta, Eşitlik (**3.3)**'teki *weight1* değeri 0.3 ve *weight2* değeri 0.7 olarak alınmıştır.

Çizelge 4.2 : Tasarlanan Öneri Modeli Sonuçları (weight1 = 0.3, weight2 = 0.7).

Tasarlanan Altyapı Modeli	Kesinlik	Anma	F-Ölçütü
DS1	0.5515	0.1751	0.2658
DS2	0.5652	0.1783	0.2711
DS3	0.5121	0.1607	0.2446
DS4	0.5157	0.1643	0.2492
DS5	0.5384	0.1659	0.2536
DS6	0.5682	0.1811	0.2747
DS7	0.5387	0.1721	0.2609
DS8	0.5301	0.1705	0.2581
DS9	0.5535	0.1756	0.2666
DS10	0.5402	0.1698	0.2585
Ortalama	0.5414	0.1713	0.2603

Çizelge **4.3**'de verilen sonuçta, Eşitlik (**3.3)**'teki *weight1* değeri 0.7 ve *weight2* değeri 0.3 olarak alınmıştır.

Çizelge 4.3 : Tasarlanan Öneri Modeli Sonuçları (weight1 = 0.7, weight2 = 0.3).

Tasarlanan Altyapı Modeli	Kesinlik	Anma	F-Ölçütü
DS1	0.6907	0.2193	0.3329
DS2	0.6473	0.2042	0.3105
DS3	0.6260	0.1964	0.2990
DS4	0.6141	0.1957	0.2968
DS5	0.6875	0.2118	0.3238
DS6	0.6123	0.1952	0.2960
DS7	0.5905	0.1887	0.2860
DS8	0.6077	0.1955	0.2959
DS9	0.6473	0.2053	0.3118
DS10	0.6777	0.2131	0.3242
Ortalama	0.6401	0.2025	0.3077

Dördüncü deneyde, tasarlanan öneri modelinin başarımını göstermek için üç deneyden farklı olarak mobil servis operatöründen elde edilen konuşma verisindeki kullanıcılar, Web sitesinden elde edilen İstanbul içindeki restoranlara giden kullanıcılarla bir kez daha rassal olarak eşleştirilerek tasarlanan öneri altyapı mimarisinin tutarlı bir şekilde çalışması gösterilmiştir. Bu deneyde önceki deneylerde eşleştirilen kullanıcılardan farklı olmak üzere kullanıcılar farklı kullanıcılarla eşleştirilmiştir.

Çizelge **4.4**'de verilen sonuçta önceki deneylere göre farklı kullanıcılar birbirleriyle eşleştirilmiş ve restoranlar arasındaki benzerlik ifadesi olan Eşitlik (**3.3)**'teki *weight1* ve *weight2* değerleri 0.5 olarak alınmıştır.

Çizelge 4.4 : Tasarlanan Öneri Modeli Sonuçları (weight1 = 0.5, weight2 = 0.5).

Tasarlanan Altyapı Modeli	Kesinlik	Anma	F-Ölçütü
DS1	0.5051	0.1603	0.2434
DS2	0.4975	0.1570	0.2387
DS3	0.4918	0.1543	0.2349
DS4	0.5157	0.1643	0.2492
DS5	0.5384	0.1659	0.2536
DS6	0.4933	0.1573	0.2385
DS7	0.5603	0.1790	0.2713
DS8	0.5043	0.1622	0.2455
DS9	0.5178	0.1643	0.2494
DS10	0.4549	0.1430	0.2176
Ortalama	0.5079	0.1608	0.2442

4.2.2 Rassal öneri modeliyle karşılaştırma

Deneysel çalışmaların ikinci kısmı için tasarlanan öneri altyapısı modelinin sonuçları, oluşturulan rassal kullanıcı modelinin sonuçları ile karşılaştırılmıştır. Konuşma verisindeki kullanıcılarla, restoranları ziyaret eden kullanıcılar eşleştirildiği için geliştirilen altyapı modeli, rassal model ile karşılaştırılmıştır. Daha öncede belirtildiği gibi tasarlanan öneri altyapı mimarisinin tutarlı bir şekilde çalışmasını göstermek için mobil servis operatöründen elde edilen konuşma verisindeki kullanıcılar, bir Web sitesinden elde edilen İstanbul içindeki restoranlara giden kullanıcılarla rassal olarak eşleştirilmiştir.

Rassal kullanıcı modelinin değerlendirilmesi için 10 farklı veri kümesi oluşturulmuş ardından rassal model sınanmıştır. 10 farklı veri kümesinin (*DSRi, i* = 1,…,10, *Çizelge **4.5***) her biri eşit sayıda kullanıcıyı içermektedir. Her bir veri kümesi konuşma verisinden rassal olarak seçilen 30 kullanıcıdan oluşmaktadır. Veri kümesinin toplamı 300 kullanıcıdan oluşmaktadır.

Her bir veri kümesi için alttaki adımlar rassal modelin başarımını ölçmek için gerçekleştirilmiştir.

İlk adımda öneri üretebilmek amacıyla, her bir kullanıcı için beş rassal kullanıcı belirlenmiştir. İkinci adımda önerilerin yapılacağı kullanıcı, yine aktif kullanıcı olarak isimlendirilmekte; aktif kullanıcının gitmediği, beş rassal kullanıcının gittiği restoranlar aktif kullanıcıya öneri olarak sunulmaktadır. Her bir aktif kullanıcı için yapılan öneri sayısı, kullanıcının test kümesindeki restoran sayısının %33'ü kadardır. Eğer önerilen bir restoran aktif kullanıcının test kümesinde bulunursa; bu restoran önerisi aktif kullanıcı için doğru bir öneri olarak belirlenmektedir.

Bütün kullanıcılar için yukarıda verilen adımlara göre öneriler yapıldıktan sonra Kesinlik, Anma, F-Ölçütü metrikleri her veri kümesi için hesaplanmaktadır. Yapılan bu deney her bir veri kümesi için on kez tekrarlanmakta ve bu on tekrarlamanın ortalama Kesinlik, Anma, F-Ölçütü sonuç ölçümleri ilgili veri kümesinin Kesinlik, Anma, F-Ölçütü değerleri olmaktadır. Rassal modelin başarımının ölçülmesi için oluşturulan, 10 veri kümesi için hesaplanan Kesinlik, Anma ve F-Ölçütü değerleri Çizelge **4.5**'deki gibi verilmiştir.

Çizelge 4.5 : Rassal Kullanıcı Modeli Sonuçları.

Rassal Kullanıcı Modeli	Kesinlik	Anma	F-Ölçütü
DSR1	0.0994	0.0324	0.0489
DSR2	0.1320	0.0420	0.0637
DSR3	0.1301	0.0416	0.0631
DSR4	0.1864	0.0598	0.0906
DSR5	0.0668	0.0212	0.0322
DSR6	0.1883	0.0602	0.0913
DSR7	0.0695	0.0221	0.0336
DSR8	0.1753	0.0571	0.0861
DSR9	0.2490	0.0859	0.1277
DSR10	0.1424	0.0452	0.0686
Ortalama	0.1439	0.0468	0.0706

Çizelge **4.1**, Çizelge **4.2**, Çizelge **4.3** ve Çizelge **4.4**'de görüldüğü gibi tasarlanan öneri altyapısının başarım ölçüm değerleri, rassal modele göre önemli ölçüde yüksek çıkmıştır. Deneysel çalışmanın gösterdiği üzere sosyal ağ ve topluluk bilgilerinin kullanılması öneri modelinin başarımını ve doğruluğunu arttırmıştır. Kullanıcıların ortam ve sosyal ağ bilgilerinin kullanılarak tasarlanan öneri altyapısının ürettiği önerilerin ortalama Kesinlik değeri %59 olarak bulunmuştur.

Tez çalışmasında geliştirilen modelin avantajları şu şekilde sıralanabilir.

- Sosyal ağ verisi kullanılarak arama uzayı küçültülmüş ve öneri üretilmesi sırasında yapılan hesaplamalar basitleştirilmiştir böylece hesap maliyeti azaltılmıştır.
- Deneysel sonuçlarda da verildiği gibi tez çalışmasında tasarlanan öneri altyapısı, rassal kullanıcı modeli ile Kesinlik, Anma ve F-Ölçütü gibi metrikler kullanılarak karşılaştırılmış ve başarımının yüksek olduğu gösterilmiştir.
- Tasarlanan öneri altyapısı, sosyal ağ verisindeki arkadaşlık bilgisini kullandığı için öneri sistemi ile kullanıcı arasındaki öneri üretme aşamasındaki iletişim oldukça kısıtlı olmaktadır. Öneri sistemi tarafından kullanıcıların herhangi bir geri bildirimine ihtiyaç duyulmamaktadır.

5. SONUÇ VE İLERİ ÇALIŞMALAR

Bu tez çalışmasında, mobil servis sağlayıcı operatörleri için sadece ortam ya da konum bilgisinin kullanıldığı öneri sistemlerine göre; hem konum ve ortam bilgisinin hem de oluşturulan sosyal ağ analizi bilgisinin kullanıldığı daha doğru sonuç veren, hızlı ve az bellek kullanımını sağlayan bir öneri sistemi için altyapı prototipi oluşturulmuştur. Oluşturulan prototip ile sosyal ağ yapısı kullanılarak, kullanıcı geri bildirimine ihtiyaç duymayan ve öneri arama uzayını küçülten bunun yanında işbirlikçi filtreleme, içerik filtreleme ve sosyal ağ analizi yöntemlerinin birleştirildiği öneri altyapısı mimarisinin oluşturulması gerçeklenmiştir. Konum ve zaman bilgileri, öneri altyapısı için kullanılan ortam bilgileri olarak tanımlanmıştır. Öneri modeli, sosyal ilişkileri ve sosyal toplulukları bulan sosyal ağ yaklaşımı ile desteklenmiştir. Sosyal ilişkilerin, işbirlikçi filtreleme gibi geleneksel öneri modelleri ile birlikte kullanılması kullanıcının ihtiyaçlarına uygun önerilerin yapılmasını sağlamıştır.

Tez çalışmasında önerilen modelde, "*Birbirleriyle sık konuşan kişiler, günlük hayatlarında da birlikte zaman geçirme eğilimi gösterirler.*" varsayımını desteklemek için sosyal ağ verisi olarak ülkemizde hizmet veren bir mobil servis sağlayıcısından elde edilmiş kişilerin gerçek konuşma verisinden oluşan veri kümesinden yararlanılmış ve birbirleriyle sık konuşan topluluklar rasgele yürüyüş algoritması ile bulunmuştur. Kişisel gizlilik ilkesi doğrultusunda konuşma verisindeki kişilerin kimlik bilgileri veri kümesi içinde şifrelenmiş olarak tutulmaktadır. Çalışmada kullanılan mekan verisi ve bu mekanları ziyaret etmiş kullanıcılardan oluşan mekan tercih bilgileri geliştirilen ayrıştırıcı ile oluşturulmuştur.

Mobil servis sağlayıcıları geliştirilen altyapı dört aşamadan oluşmaktadır. Birinci aşamada kullanıcı mobil telefon arayüzü ile öneri isteğinde bulunmakta, ikinci aşamada kullanıcının konumu Küresel Yer Belirleme Sistemi (GPS) ile belirlenmekte, üçüncü aşamada içerik filtreleme, işbirlikçi filtreleme ve sosyal ağ analizi yöntemleri kullanılarak mekan önerileri üretilmekte, son aşamada ise kullanıcılara önerilen mekanlar çeşitli mobil interaktif kanalları kullanılarak sunulmaktadır. Geliştirilen altyapı mimarisinin kullanımını modellemek için mobil telefon simülatörü tasarlanmış ve kullanıcılara mobil telefon arayüzü aracılığı ile mekan önerileri sunulmuştur.

Deneysel sonuçlarda da verildiği gibi tez çalışmasında tasarlanan öneri altyapısı, rassal kullanıcı modeli ile Kesinlik, Anma ve F-Ölçütü gibi metrikler kullanılarak karşılaştırılmış ve başarımının yüksek olduğu gösterilmiştir. Deneysel çalışmada elde edilen ön sonuçlar umut verici olarak değerlendirilmektedir.

Geliştirilen altyapı mimarisi esnek, genişletilebilir, yeniden kullanılabilir ve farklı alanlardaki uygulamalara kolay adapte olması özelliğiyle değişik ürünlerin önerilerinde de kullanılacak modele kolaylıkla dönüştürülebilir. İlerleyen çalışmalarda da tez çalışmasında geliştirilen altyapı mimarisinin otel, sinema, hastane önerilerinde de kullanılmasına çalışılacaktır. Mevcut literatür araştırmaları incelendiğinde tez çalışmasında gerçeklenen öneri altyapısı, mobil ortamlarda sosyal ağ verisi olarak gerçek konuşma verisinin kullanıldığı bir ilk çalışmayı oluşturmaktadır.

KAYNAKLAR

[1] **Golbeck, J.,** 2006: Generating Predictive Movie Recommendations from Trust in Social Networks, *Proceedings of iTrust International Conference,* Pisa, Italy, pp. 93-104.

[2] **Latapy, M., and Pons, P.,** 2005: Computing communities in large networks using random walks, *Proceedings of 20th International Symposium on Computer and Information Sciences,* Istanbul, Turkey.

[3] **Newman, M. E. J.,** 2003. The structure and function of complex networks, *SIAM REVIEW,* 45-167.

[4] **Bearman, P. S., and Moody, J.,** 2002: Chains of affection: The structure of adolescent romantic and sexual networks, Preprint, Department of Sociology, Columbia University.

[5] **Jeong, H., Mason, S., Barabasi, A., and Oltvai, Z. N.,** 2001: Lethality and centrality in protein networks, *Nature,* 41-42.

[6] **Chen, Q., Chang, H., Govindan, R., Jamin, S., Shenker, S. J., and Willinger, W.,** 2002: The origin of power laws in Internet topologies revisited, *Proceedings of 21th Annual Joint Conference of the IEEE Computer and Communications Societies,* IEEE Computer Society.

[7] **Ebel, H., Mielsch, L., and Bornholdt, S.,** 2002: Scale-free topology of e-mail networks, *Phys. Rev. E,* 035103.

[8] **Watts, D. J., and Strogatz, S. H.,** 1998: Collective Dynamics of small world networks, *Nature,* 440-442.

[9] **Redner, S.,** 1998: How popular is your paper? An empirical study of the citation distribution, *Eur. Phys. J. B,* 131-134.

[10] **Pazzani, M. S.,** 1999: A Framework for Collaborative, Content-Based and Demographic Filtering, *Artificial Intelligence Review*, Vol. 13, pp 393-408.

[11] **Weiyang, L., Alvarez, S., and Ruiz,C.,** 2002. Efficient Adaptive-Support Association Rule Mining for Recommender Systems, Data Mining and Knowledge Discovery, Springer Netherlands, ISSN: 1384-5810, Volume 6, Number 1 / January, 2002, pp. 83-105

[12] **Shani, G., Heckerman, D., and Brafman, R.,** 2005. An MDP-Based Recommender System, Journal of Machine Learning Research, pp. 1265-1295

[13] **Yan, T., Jacopsen, W., Garcia-Molina, M., and Dayal, H. U.,** 1996. From user access patterns to dynamic hypertext linking, The Fifth World Wide Web Conference (WWW5), Paris, France, May 6-10, pp. 1007-1014

[14] **Mobasher, B., Dai, H., Luo, T., and Nakagawa, M.,** 2002.Using Sequential and Non-Sequential Patterns in Predictive Web Usage Mining Tasks, School of Computer Science, Telecommunication, and Information Systems DePaul University, Chicago, USA

[15] **Herlocker J. L.,** 2004: Evaluating collaborative filtering recommender systems, *In ACM Transactions on Information Systems*, Volume 22, No. 1, 2004, pp 5-53.

[16] **Billsus, D., and Pazzani, M.,** 1998. Learning Collaborative Filters. *Proceedings of ICML'98*, 46-53. Morgan Kaufman Eds.

[17] **Salton, G., and McGill, M.,** 1983: Introduction to Modern Information Retrieval, McGraw-Hill, New York, NY, USA.

[18] **Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J.,** 1994: Grouplens: an open architecture for collaborative filtering of netnews, *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp. 175–186, New York, NY, USA.

[19] **Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl, J.,** 2000: Analysis of recommendation algorithms for E-commerce, *Proceedings of the ACM E-Commerce*, pp. 158–167, Minneapolis, Minn, USA.

[20] **Breese, J. S., Heckerman, D., and Kadie, C.,** 1998: Empirical Analysis of Predictive Algorithms for Collaborative Filtering, Technical Report MSR-TR-98-12, Microsoft Research, Redmond, USA

[21] **Sobecki, J.,** 2006: Implementations of Web-based Recommender Systems Using Hybrid Methods, *International Journal of Computer Science & Applications*, Vol. 3, pp 52-64.

[22] **Turnbull, D.,** 1998: Augmenting Information Seeking on the World Wide Web Using Collaborative Filtering Techniques, from http://donturn.fis.utoronto.ca/research/augmentis.htn

[23] **Turnbull, D.,** 1997: KMDI Final Summary: Collaborative Filtering, from http://donturn.fis.utoronto.ca/research/kmdi-cf.html

[24] **Wittenburg, K., Duco, D., Will, H., and Larry, S.,** 1998: Group Asynchronous Browsing on the World Wide Web, from http://www.w3.org/Conferences/WWW4/Papers/98/

[25] **Lee, H. J., Choi, J. Y., and Park, S. J.,** 2005: Context-Aware Recommendations on the Mobile Web, OTM Workshops, LNCS 3762, pp. 142-151

[26] **Dey, A. K., and Abowd, G. D.,** 2000: Towards a Better Understanding of Context and Context-Awareness. In: CHI 2000 Workshop on The What, Who, Where, When, Why and How of Context-awareness, the Hague, the Netherlands

[27] **Tung, H., and Soo, V.,** 2004: A Personalized Restaurant Recommender Agent for Mobile E-Service, *Proceedings of IEEE International Conference on e-Technology,e-Commerce and e-Service*

[28] **Park, M. H., Hong, J. H., and Cho, S. B.,** 2007: Location-Based Recommendation System Using Bayesian User's Preference Model in Mobile Devices, UIC, LNCS 4611, pp. 1130-1139.

[29] **Cheverst, K., Davies, N., Mitchell, K., and Friday, A.,** 2000: Experiences of Developing and Deploying a Context-Aware Tourist Guide: The GUIDE Project. In: MOBICOM, Boston, MA, USA

[30] **Tewari, G., Youll, J., and Maes, P.,** 2002: Personalized Location-based Brokering using an Agentbased Intermediary Architecture. In: Decision Support Systems, 34 p.p 127-137

[31] **Yuan, S. T., and Tsao, Y. W.,** 2003: A Recommendation Mechanism for Contextualized Mobile Advertising. In: Expert Systems with Applications, 24 399-414

[32] **Pousman, Z., Iachello, G., Fithian, R., Moghazy, J., and Stasko, J.,** 2004: Design Iteration for a Location-aware Event Planner. In: Personal and Ubiquitous Computing, 8 117-125

[33] **Hayes, C., and Cunningham, R**., 2004: Context Boosting Collaborative Recommendations. In: Knowledge-based Systems, 17 131-138

[34] **Setten, M. C., Pokraev, S., and Koolwaaij, J.,** 2004: Context-Aware Recommendations in the Mobile Tourist Application COMPASS. In Nejdl, W. and P.D. Bra, Editors (eds) AH 2004, LNCS 3137, p. 235-244

[35] **Ricci, F., and Nguyen, Q. N.,** 2005: Critique-Based Mobile Recommender Systems, *ÖGAI Journal*, Vol. 24, No. 4

[36] **Terveen, L. G., and Hill, W. C.,** 2001: Beyond Recommender Systems: Helping People Help Each Other, in Caroll, J. (ed.), HCI In The New Millennium. Addison-Wesley

[37] **Liu, H., and Maes, P.,** 2005: InterestMap: Harvesting Social Network Profiles for Recommendations, *IUI*, San Diego, California, USA

[38] **Golbeck, J.,** 2006: Generating Predictive Movie Recommendations from Trust in Social Networks, Proceedings of Semantic Network Analysis Workshop, Galway, Ireland.

[39] **Url-1 <***http://jericho.htmlparser.net/docs/javadoc/index.html*>, alındığı tarih 15.04.2010.

[40] **Url-2 <***http://en.wikipedia.org/wiki/Web_service*>, alındığı tarih 15.04.2010.

[41] **Url-3 <***http://www.w3schools.com/webservices/default.asp*>, alındığı tarih 15.04.2010.

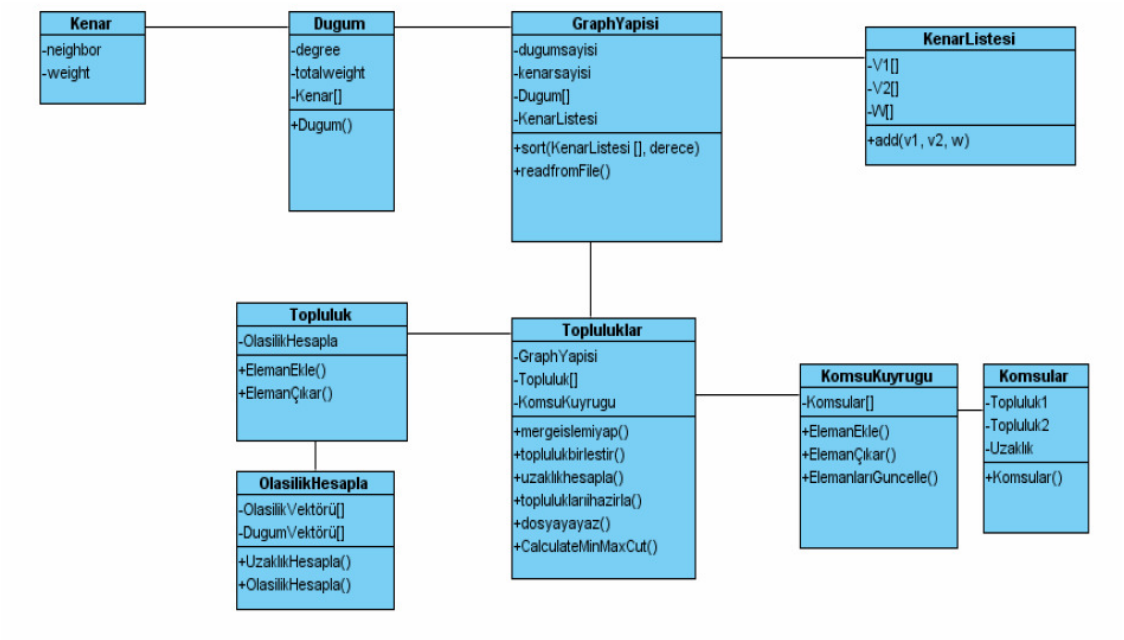
[42] **Url-4 <**http://developers.sun.com/mobility/apis/articles/wsa/>, alındığı tarih 15.04.2010.

EKLER

EK A.1 : Rasgele Yürüyüş Algoritması Sınıf Diyagramı

EK A.1

Sınıf Diyagramı:



Şekil A.1 : Sınıf Diyagramı

ToplulukBelirleme: Dugum, GraphYapisi, Kenar, KenarListesi, KomsuKuyrugu, Komsular, Main, OlasilikHesapla, Topluluk ve Topluluklar sınıflarını içerir.

Dugum sınıfı: Her bir düğüm yapısının kurulduğu yapıdır. Düğümün bağlı bulunduğu kenarları, derecesini ve toplam ağırlığını içermektedir.

GraphYapisi sınıfı: Dosyadaki düğümlerin ve ağırlıkların okunması sonucunda kurulan çizge yapısını temsil etmektedir. Toplam düğüm sayısını, toplam kenar sayısını, düğüm ve kenar listesi sınıflarını içermektedir.

Kenar sınıfı: Çizgede verilen bir düğüm için bir kenarı temsil etmektedir. Ağırlık ve kenarın karşısındaki komşu düğümü eleman olarak içermektedir.

KenarListesi sınıfı: Dosyadaki düğüm çiftlerinin ve ağırlıkların okunması sonucunda kurulan kenarların listesini içermektedir. *V1, V2, W* gibi elemanları bulunmaktadır.

KomsuKuyrugu sınıfı: Kuyruk Heap ağacı şeklinde gerçeklenmiştir. Kuyruğa ekleme, çıkarma ve güncelleme metotlarının gerçekleştirildiği sınıftır. Komşular tipinden elemanları içinde bulundurmaktadır. Her bir eleman iki topluluk çiftlerinden ve bu topluluk çiftinin arasındaki uzaklıktan oluşmaktadır. Topluluklar birleştirilip yeni topluluklar oluşturulurken kuyruktaki elemanlar güncellenmektedir. Güncelleme işlemini ve elemanların kuyrukta bulunma sırasını uzaklık değeri belirlemektedir.

OlasilikHesapla sınıfı: Her bir topluluğun olasılık vektörünün hesaplandığı ve bu hesaplama sırasında gezilen düğümlerin listesinin tutulduğu sınıftır.

Topluluk sınıfı: Her bir topluluk yapısını temsil etmektedir. Olasılık değerlerinin hesaplanması için içinde OlasilikHesapla tipinde bir nesne bulunmaktadır. Bu nesne sayesinde ilgili topluluğun olasılık vektörü ve olasılık vektörünün hesaplanması sırasında gezilen düğümlerin vektörü tutulmaktadır.

Topluluklar sınıfı: Bütün toplulukların kapsandığı sınıftır. Programın başlangıç aşamasında oluşturulur. Bütün topluluk yapılarının kurulmasından, bu topluluk yapılarının birleştirilmesinden ve dosyaya yazılmasından sorumludur.

Main sınıfı: *main* metodunun bulunduğu sınıftır. Topluluklar tipinden topluluklar sınıfı oluşturulmaktadır. Ardından oluşturulan her topluluğun olasılık değerleri ve toplulukların aralarındaki uzaklığın değeri hesaplanmakta ve son olarak oluşturulan topluluklar aşağıdan yukarıya hiyerarşik demetleme algoritması ile demetlenerek topluluklar bulunmaktadır. Sonlanma kriteri olarak kuyruktaki eleman sayısının bitmesi belirlenmiştir.

ÖZGEÇMİŞ



Alper ÖZCAN 1984'de Antalya'da doğdu. 2002 yılında Eskişehir Fen Lisesi'nden mezun oldu. 2002 yılı ÖSS sonuçlarına göre; İTÜ Bilgisayar Mühendisliğinde okumaya hak kazanmıştır. Üniversite öğrenimi boyunca İTÜ Geliştirme Vakfından Onur Bursu alarak 2007 senesinde Bilgisayar Mühendisliği Bölümünden 3.64/4.00 ortalama ile 4. olarak mezun olmuştur. Lisans bitirme ödevi olarak "FloatBoost Algoritması ile Resimlerde Yüz Yerinin Bulunması" konusunu almış ve bu çalışmada resimlerdeki insan yüzlerinin yeri FloatBoost algoritması ile bulunması sağlanmıştır. 2004 yılında başlamış olduğu İşletme Mühendisliği Çift Anadal Programından 2008 yılında mezun olarak İşletme Mühendisi ünvanı almaya hak kazanmıştır. 2007 yılı Eylül ayında Bilgisayar Mühendisliği Yüksek Lisans programına kaydolmuştur.
Yüksek Lisans Öğrenim hayatı sırasında sürdürdüğü iş yaşamında; 2008-2009 yılları arasında TÜBİTAK Marmara Araştırma Merkezi'nde araştırmacı olarak görev almış ve Denizaltı Taktik Simülatörü Projesinde yazılım mühendisi olarak uygulama geliştirmiştir. 2009'dan bugüne de Turkcell Teknoloji AR-GE şirketinde müşteri yönetimi yazılım geliştirme ekibinde görevini sürdürmektedir.

Ad Soyad: Alper ÖZCAN

Doğum Yeri ve Tarihi: Korkuteli,ANTALYA, 13.09.1984

Adres: Emniyetevleri Mah. Çelebi Mehmet Sokak 8/26 4.Levent / İstanbul

Lisans Üniversite: İstanbul Teknik Üniversitesi

Yayın Listesi:

▪ **Ozcan A**., Oguducu G. S., 2010: A Recommendation Framework For Mobile Phones Based On Social Network Data, *11th ACIS International Conference on SNPD 2010*, June 9-11, 2010 London, United Kingdom.