

PAPER DETAILS

TITLE: BIRLIKTELIK KURALLARI ILE MEKÂNSAL-ZAMANSAL VERİ MADENCİLİĞİ

AUTHORS: Muhammed ALAEDDINOĞLU,Tolga AYDIN,Deniz DAL

PAGES: 191-212

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/68428>

BİRLİKTELİK KURALLARI İLE MEKÂNSAL-ZAMANSAL VERİ MADENCİLİĞİ

SPATIO-TEMPORAL DATA MINING WITH ASSOCIATION RULES

**Muhammed Fatih ALAEDDİNOĞLU^{1*}, Tolga AYDIN² ve Deniz
DAL²**

*¹Bayburt Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü,
Bayburt*

*²Atatürk Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü,
Erzurum*

Geliş Tarihi: 6 Nisan 2012

Kabul Tarihi: 25 Temmuz 2012

ÖZET

İnsanlar için stratejik anlamda karar vermek önemlidir. Karar verebilmek için ise kavramlar ile alakalı geçmişe dayalı bilgi birikimine sahip olmak gerekir. Genellikle bu birikime sahip insanları yani uzmanları bulmak zor ve maliyetli olmaktadır. Gelişen bilgi teknolojileri ve Veri Tabanı Yönetim Sistemlerindeki (VTYS) yenilikler eldeki verilerden stratejik karar için yararlı bilgiyi çıkarmayı mümkün kılmaktadır. Uzmanlara bağlılığı azaltmak ve bilgi çıkarımını otomatik olarak yapabilmek için farklı yöntemler geliştirilmiştir. Bu yöntemlerden birisi de veri madenciliği yöntemidir. Veri madenciliği yöntemi de kendi arasında sınıflandırma, ilişki öğrenme (Birliktelik kuralları öğrenme), kümeleme gibi alt bölümlerden oluşmaktadır. Birliktelik kuralları öğrenme aşaması için literatürde en yaygın kullanılan algoritma Apriori algoritmasıdır (Agrawal ve Srikant, 1994). Klasik veri madenciliği alanında yaygın kullanıma sahip olan Apriori algoritması, zaman ve mekâna bağlı olarak değişen veri kümelerinde çok fazla kullanılmamaktadır. Zaman ve mekâna bağlı veri kümelerinin klasik veri kümelerinden farkı, verilerin klasik veri özelliklerine zaman ve konum özelliklerinin de eklenmesidir. Bu çalışmada Apriori algoritması, zamansal-mekânsal veri madenciliği alanına uyarlanarak, Van Gölü'ne ait zamansal-mekânsal veri kümesi üzerinde uygulanmıştır.

Anahtar Kelimeler: Veri Madenciliği, Birliktelik Kuralı, Apriori Algoritması, Zamansal-Mekânsal Veri

* Sorumlu yazar: falaeddinoglu@bayburt.edu.tr

ABSTRACT

It is important to give strategic decisions for the human-beings. On the other hand, it is required to have past knowledge-base related to the concepts to give decisions. It is generally too hard and expensive to find such knowledgeable experts. Developing information technologies and the improvements in the database management systems make it possible to extract useful information from knowledge in hand for the strategic decisions. Different methods have been developed to reduce expert-dependency and to extract information automatically. Data mining is one of such methods. Data mining consists of sub-disciplines such as classification, association learning (association rules learning) and clustering. Apriori algorithm (Agrawal ve Srikant, 1994) is the commonly used association rules learning algorithm in the literature. Although Apriori algorithm has a broad usage in the classical data mining area, it is rarely used in spatio-temporal data sets. Spatio-temporal data sets are different from classical data sets in the sense that time and location properties are embedded to the classical data properties. In this study, Apriori algorithm is adapted to the spatio-temporal data mining area and used on a spatio-temporal data set belonging to Van Lake.

Keywords: Data Mining, Association Rules, Apriori Algorithm, Spatio-Temporal Data

1. GİRİŞ

İnsanlar, tarih boyunca geçmişe dayalı tecrübelerini kullanarak geleceğe yönelik tahminler ve çıkarımlar yapmışlardır. Günümüzde, geleceğe yönelik yapılan öngörüler çok fazla önem kazanmıştır. Bilhassa yatırım amaçlı yapılan çalışmalarda geçmişin çok fazla önemi vardır. Her geçen gün işletmeler ve devlet kurumları veri tabanı sistemine daha fazla yatırım yapmakta ve daha fazla oranda veriyi bu sistemlerde depolamaktadır (Özçakır ve Çamurcu, 2007). Gelişen teknolojiler sayesinde geçmişe dayalı veriler bilgisayar ortamında işlenerek çok daha hızlı ve etkin tahmin ve çıkarımlar yapmak mümkün olmaktadır.

Geleceğe yönelik tahminlerde bulunmak için bilgisayar teknolojileri oldukça yaygın kullanılmaktadır. Yapay zekâ teknikleri, uzman sistemler ve veri madenciliği teknikleri kullanılarak yazılan programlar sayesinde geçmişte kayıt altına alınmış veriler

değerlendirip bu veriler doğrultusunda sonraki oluşacak olan durumlar tahmin edilmeye çalışılır. Bu çalışmada veri madenciliği, yani veritabanından bilgi keşif süreci tekniği (Han ve Kamber, 2006) kullanılmıştır.

Veri madenciliği, büyük ölçekli veriler arasından “değerli olan” bilgiyi elde etme işidir. Veri madenciliği ile veriler arasındaki ilişkileri ortaya koymak ve gerektiğinde ileriye yönelik kestirimlerde bulunmak mümkündür. Bu yöntem kendi başına bir çözüm değil, çözüme ulaşmak için verilecek karar sürecini destekleyen, problemi çözmek için gerekli bilgileri sağlamaya yarayan bir araçtır. Veri madenciliğini kendi içinde üç bölüme ayırılır. Bunlar; sınıflama, kümeleme, birliktelik kuralları öğrenmedir (Karabatak ve İnce, 2010). Bu çalışmada bu tekniklerden birliktelik kuralları öğrenme alt tekniği kullanılmıştır.

Birliktelik kuralları, veriler arasındaki ilişkileri ortaya çıkararak kurallar oluşturur. Bu kurallara bağlı olarak daha sonra verilerin birlikte olup olmama durumları tahmin edilir. Örnek olarak; markette yapılan bir alışverişte peynir alan müşterilerin belirli bir oranının aynı zamanda soda da aldığı tespit edildiği takdirde peynir ve soda ürünleri arasındaki birliktelik ilişkisi ortaya çıkmış olur. Yani müşterilerin alışveriş hareketlerinden genel kurallar çıkarılarak ürünlerin marketteki raf düzeni belirlenebilir, kampanyalar oluşturabilir veya ürün gereksinimleri belirlenerek depo stokları ayarlanabilir. Bütün bunlar market ürün satışlarının artması ile sonuçlanır ki, bu da birliktelik kuralı öğrenmenin elle tutulur faydalarındandır. Milyonlarca veri üzerinde birliktelik kuralı öğrenme gibi veri madenciliği teknikleri uygulandığında, kullanılacak algoritmalar hızlı olmalıdır (Agrawal ve Srikant, 1995).

Günümüzde bir çok alan ile alakalı veriler kayıt altına alınmaktadır. Kayıt altına alınan bu verilerden bazıları özellikle coğrafi veriler tek veya iki boyutlu olarak ifade edilememektedir. Örnek olarak bir şehirdeki yapıların kapladığı alanın zamana bağlı olarak değişmesi klasik veri türleriyle modellenememektedir. Bunun sonucu olarak yeni veri türlerinin modellenmesi ihtiyacı ortaya çıkmıştır. Veri Tabanı Yönetim Sistemleri (VTYS) yazılımları bu ihtiyacı gidermek adına mekânsal verileri ifade edebilmek için

“geography” veya “geometry” adlarında yeni veri tipleri oluşturmuşlardır (Sack, 2010).

Birlikte kuralları öğrenme aşaması için literatürde en yaygın kullanılan algoritma Apriori algoritmasıdır (Agrawal ve Srikant, 1994). Klasik veri madenciliği alanında yaygın kullanıma sahip olan Apriori algoritması, zaman ve mekâna bağlı olarak değişen veri kümelerinde çok fazla kullanılmamaktadır. Zaman ve mekâna bağlı veri kümelerinin klasik veri kümelerinden farkı, verilerin klasik veri özelliklerine zaman ve konum özelliklerinin de eklenmesidir. Bu çalışmada Apriori algoritması, zamansal-mekânsal veri madenciliği alanına uyarlanarak, deneysel bir uygulama üzerinde anlatılacaktır.

2. MATERYAL VE METOT

Geçmişte kayıt altına alınmış mekânsal (konumsal) verilerin zaman içerisindeki değişiminde rol oynayan faktörler arasındaki ilişkileri ortaya koyacak kurallar oluşturulması esastır. Oluşturulan kurallar sayesinde konumsal bilgilerin zamanla nasıl değişeceği tahmin edilir. Bu çalışmada kuralların elde edilmesi için birlikte kurallarından Apriori Algoritması kullanılmıştır. İncelenen veriler zamansal-mekânsal veriler olduğundan dolayı Apriori Algoritması bu bağlamda uyarlanarak zamansal-mekânsal bir veri madenciliği yöntemi kullanılmıştır.

2.1. Veri Madenciliği

Veri madenciliği, büyük ölçekli veriler arasından “değerli olan” bilgileri elde etme işidir. Veri madenciliği ile veriler arasındaki ilişkileri ortaya koymak ve gerektiğinde ileriye yönelik kestirimlerde bulunmak mümkündür. Bu yöntem karar destek sistemleri için önemli bir yere sahiptir. Veri madenciliği, kendi başına bir çözüm değil çözüme ulaşmak için verilecek karar sürecini destekleyen, problemi çözmek için gerekli bilgileri sağlamaya yarayan bir araçtır.

2.1.1. Veri Madenciliği Uygulama Alanları

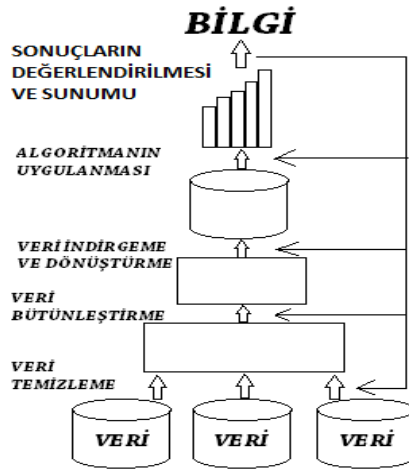
Uygulama alanlarına ait bazı örnekler aşağıda verilmiştir:

- Pazar araştırması; hedef pazar, müşteriler arası benzerliklerin saptanması, sepet analizi, çapraz pazar incelemesi

- Risk analizi; kalite kontrol, rekabet analizi, sahtekârlıkların saptanması
- Belgeler arası benzerlik; e-postalar, haber kümeleri benzerliklerinin saptanması
- Müşteri kredi risk araştırmaları
- Yaban hayatı yönetimi; popülasyonun yörünge izleri izlenerek hayvan göç modellerini ortaya koyma
- Kirlilik; kayıtlı duman iz hareketleri izlenerek, hava akışı modelleri üzerinde çalışarak kirlilik kaynağına ulaşabilme
- Sensörler; fiziksel alanlarda çevresel izlenimler
- Coğrafi Bilgi Sistemi; deprem, su taşmaları, arazi oluşumları, kent yerleşim planı, yatırım amaçlı alanların belirlenmesi gibi coğrafi çalışmalar

2.1.2. Veri Madenciliği Aşamaları

Bir veri kümesine veri madenciliği tekniğinin uygulanması için verilerin kullanılabilir hale dönüştürülmesi gerekmektedir. Bu sebeple verilerin Şekil 1' deki gibi bazı alt aşamalardan geçmesi gerekmektedir (Han, 2000).



Şekil 1. Veri madenciliği Aşamaları

2.1.2.1. Veri Temizleme

Üzerinde çalışılacak olan verilerden bazıları işlenmeye uygun olmayabilir. Böyle verilere gürültü denir. Bu durumlarda gürültülü verilerin temizlenmesi gerekmektedir. Bu işlem için aşağıdaki yöntemlerden biri veya daha fazlası uygulanabilir (Han, 2000).

- Eksik değer içeren kayıtlar veri kümesinden atılabilir.
- Kayıp değerlerin yerine genel bir sabit kullanılabilir.
- Ortalama bir değer kullanılabilir.
- Regresyon veya karar ağacı kullanılarak veri tahmin edilebilir.

2.1.2.2. Veri Bütünleştirme

Farklı veri tabanlarından ya da veri kaynaklarından alınan verilerin birlikte değerlendirmeye alınabilmesi için farklı türdeki verilerin tek türe dönüştürülmesi işlemine denir (Özkan, 2008).

2.1.2.3. Veri İndirgeme

Birbirine paralel, örtüşen, yakın verilen azaltılması işlemidir. Aynı zamanda genelleme veya veri sıkıştırma işlemleri yapılarak da veri indirgenebilir. Ancak sonucu belli bir oranın üzerinde etkileyecek olan veri indirgemelerinden kaçınılmalıdır (Özkan, 2008).

2.1.2.4. Veri Dönüştürme

Bazı verileri, uygulamada direkt kullanmak uygun olmayabilir. Böyle durumlarda veriyi ya kullanabileceğimiz bir yapıya ya da boyuta dönüştürmemiz gerekebilir. “Double” veri türüne sahip veriyi “integer” veri türüne dönüştürmek veya çok büyük bir spektruma sahip verileri [0,1] aralığına toplama işlemi veri dönüştürme aşamasına örnek olarak verilebilir.

2.1.2.5. Veri Madenciliği Algoritmalarını Uygulama

Önceki adımlarda belirtilen işlemlerden sonra uygun olan veri madenciliği tekniğini uygulamaktır.

2.1.2.6. Sonuçların Değerlendirilmesi ve Sunumu

Veri madenciliğinin uygulanması sonucu ortaya çıkan verilerin uzman kişilerce değerlendirilmesidir. Eğer sonuçlar uygunsa ortaya çıkan sonuçların kullanılması, aksi takdirde uygun bir adıma geri dönülerek işlemlerin yeniden yapılması aşamasıdır.

2.1.3. Veri Madenciliği Yöntemleri

Günümüzde yaygın olarak kullanılmaya başlayan bu tekniğin uygulanması için farklı yöntemler geliştirilmiştir. Çalışılan alana ve amaca göre bu yöntemlerden biri veya birkaçı seçilebilir. Genel olarak üç sınıfa ayrılmıştır. Ancak çalışmamızda birliktelik kuralları yöntemini kullandığımız için sadece bu yöntem ayrıntılı olarak anlatılacaktır.

2.1.3.1. Sınıflandırma

Verilerin içerdiği ortak özelliklere göre sınıflandırıldığı yöntemidir. Bu yöntem sayesinde yeni öğelerin özelliklerine bakılarak veri sınıfı hakkında tahminde bulunulabilir.

2.1.3.2. Kümeleme

Birbiriyle benzerlik veya yakınlık gösteren verilerin gruplandırılması işlemidir. Genellikle bu benzerlik, konumların benzerliği veya yakınlığı ile ilintilidir.

2.1.3.3. Birliktelik Kuralları

Olayların birlikte gerçekleşme durumlarını inceleyen yöntemdir. Birliktelik kuralları (*Association rules*), olayların birbirlerine bağlı olarak gerçekleşme durumlarını olasılık terimleriyle ortaya koyar. Bu yöntemin temel amacı, veriler arasındaki ilginç ilişkileri ortaya çıkarmaktır. İstatistiksel olarak yapılan çalışmalar sonucunda ortaya çıkan sonuçlar genellikle çok şaşırtıcı olmazken, birliktelik kuralları ile çok daha ilginç çıkarımlar yapmak mümkün olmaktadır (Şentürk, 2006). Örneğin; çocuk bezi satın alan müşterilerin büyük bir olasılıkla beraberinde mama da satın aldığını saptamak çok da zor bir iş değildir. Ancak çocuk bezi satın alan müşterilerin çoğunlukla bira da aldığını öngörmek oldukça zor bir iştir (Şentürk, 2006). Satın alma olayındaki bu beraberlik ilişkisini, birliktelik kuralları analizi ile ortaya çıkarmak mümkündür. Birliktelik kurallarının matematiksel modeli Agrawal, Imielinski ve Swami tarafından 1993 yılında sunulmuştur (Agrawal vd., 1993). Bu modelde, $I = \{i_1, i_2, \dots, i_m\}$ kümesine "nesneler" adı verilmektedir. D , veri bütünlüğündeki tüm hareketleri, T ise ürünlerin her bir hareketini simgeler. TID ise, her harekete ait olan tek belirteçtir. Birliktelik kuralı şu şekilde tanımlanabilir:

$$A_1, A_2, \dots, A_m \rightarrow B_1, B_2, \dots, B_n$$

Bu ifadede yer alan, A_i ve B_j yapılan iş veya nesnelerdir. Bu kural, genellikle " $A_1, A_2, \dots, A_m$ " iş veya nesneleri meydana geldiğinde, sık olarak " $B_1, B_2, \dots, B_n$ " iş veya nesnelerinin de aynı olay veya hareket içinde yer aldığını belirtir (Zhu, 1998).

Birliktelik kuralları, kullanıcı tarafından minimum değeri belirlenmiş destek ve güvenilirlik eşik değerlerini sağlayacak biçimde üretilir.

A ve B nesne kümelerinin birliktelik kuralı " $A \rightarrow B$ " olarak gösterilirse, destek aşağıdaki gibi tanımlanır:

$$\text{destek } (A \rightarrow B) = (\text{D içinde A ve B'nin beraber bulunduğu kayıt sayısı}) / |D|$$

Yani; destek, D ile ifade edilen tüm hareketler içinde ilgili nesne kümesini ($A \cup B$) içeren hareketlerin oranıdır.

$A \rightarrow B$ birliktelik kuralının güven değeri ise, A'yı içeren hareketlerin B'yi de içermesi olasılığı olarak tanımlanır. Örneğin, $A \rightarrow B$ kuralı % 85 güvenilirliğe sahip ise, A'yı içeren nesne kümelerinin % 85'i B'yi de içermektedir.

$$\text{güven } (A \rightarrow B) = (\text{D içinde A ve B'nin beraber bulunduğu kayıt sayısı}) / (\text{D içinde sadece A'nın bulunduğu kayıt sayısı})$$

Birliktelik kurallarına ilişkin olarak geliştirilen bazı algoritmalar şunlardır:

AIS (Agrawal vd., 1993), SETM (Houtsma ve Swami, 1995), Apriori (Agrawal ve Srikant, 1994), Partition (Savasere vd., 1995), RARM - Rapid Association Rule Mining (Das vd., 2001), CHARM (Zaki ve Hsiao, 2002).

Bu algoritmalar, kullanım yerlerine göre birbirlerine üstünlük gösterebilmektedirler. Bu makalede, çalışmamıza en uygun olduğunu düşündüğümüz Apriori Algoritması kullanılmıştır.

2.1.3.3.1. Apriori Algoritması

Apriori Algoritmasının ismi, bilgileri sürekli bir önceki adımdan aldığı için "önceki" (prior) anlamına gelen "Apriori" den

gelmektedir (Agrawal ve Srikant, 1994). Bu algoritma, temelinde iteratif (tekrarlayan) bir niteliğe sahiptir (Han ve Kamber, 2006) ve hareket bilgileri içeren veritabanlarında sık geçen öge kümelerinin keşfedilmesinde kullanılır. Apriori algoritmasının özüne göre, herhangi bir k-öge kümesi (k adet elemana sahip öge kümesi) minimum destek ölçütünü sağlıyorsa, bu kümenin alt kümeleri de minimum destek ölçütünü sağlar.

k-öge (k adet elemana sahip öge kümesi) kümesi c ile ifade edilirse, öğeleri (ürünler, nesneler) $c[1], c[2], c[3], \dots, c[k]$ şeklinde gösterilir ve $c[1] < c[2] < c[3] < \dots < c[k]$ olacak şekilde küçükten büyüğe doğru sıralıdır (Agrawal ve Srikant, 1994). Her öge kümesine destek ölçütünü tutmak üzere bir sayaç değişkeni eklenmiştir ve sayaç değişkeni öge kümesi ilk oluşturulduğunda sıfırlanır. Sık geçen öge kümeleri "L" karakteri ile, aday öge kümeleri ise "C" karakteri ile gösterilir (Sever ve Oğuz, 2002). Algoritmanın genel yapısı aşağıda yalancı kod şeklinde verilmiştir:

```

L1 = {sık geçen 1-öge kümesi};
for (k=2; Lk-1 ≠ ∅; k++) do begin
    Ck = apriori-gen (Lk-1); // Yeni adaylar
    for all t in D do begin // transactions-hareketler
        Ct = subset (Ck, t); // Adaylar t içindedir
        for all c in Ct do // candidates - adaylar
            c.count++;
    end
    Lk = {c ∈ Ck | c.count ≥ minsup}
end
Answer = Uk Lk;

```

Algoritmanın işleyişini bir örnekle açıklayalım (Minimum destek parametresini 50%, sık nesne (ürün) kümelerinden birliktelik kuralları oluşturma aşamasında kullanılacak minimum güven parametresini de 50% alalım). Tablo 1'deki her bir işlem bir alışveriş fişini ifade etsin.

Tablo 1. Alışveriş İşlemleri (Fişleri)

İşlem ID	Öğeler
1	Ayakkabı, Gömlek, Ceket, Pantolon, Çorap
2	Ayakkabı, Ceket, Mont
3	Ayakkabı, Pantolon, Çorap, Mont
4	Gömlek, Süveter, Tişört

Tablo 2’de 1-öğeli kümelerin destek değerleri verilmiştir. Tablo 2’deki 1-öğeli kümelerden minimum destek değeri 50%’den büyük olanlar ise Tablo 3’de verilmiştir.

Tablo 2. 1-öğeli Kümelerin Destek Değerleri

Öğeler	Destek Değerleri
Ayakkabı	75%
Gömlek	50%
Ceket	50%
Pantolon	50%
Mont	50%
Çorap	50%
Tişört	25%
Süveter	25%

Tablo 3. Destek Değeri Minimum Destek Değerinden Büyük 1-öğeli Kümeler

Öğeler	Destek Değerleri
Ayakkabı	75%
Gömlek	50%
Ceket	50%
Pantolon	50%
Mont	50%

Tablo 4’de 2-öğeli kümelerin destek değerleri verilmiştir. Tablo 4’deki 2-öğeli kümelerden minimum destek değeri 50%’den büyük olanlar ise Tablo 5’de verilmiştir.

Tablo 4. 2-öğeli Kümelerin Destek Değerleri

Öğeler	Destek Değerleri
Ayakkabı, Gömlek	25%
Ayakkabı, Çorap	50%
Ayakkabı, Ceket	50%
Ayakkabı, Pantolon	50%
Ayakkabı, Mont	50%
Gömlek, Ceket	25%
Gömlek, Çorap	25%
Gömlek, Pantolon	25%
Gömlek, Süveter	25%
Gömlek, Tişört	25%
Ceket, Pantolon	25%
Ceket, Çorap	25%
Ceket, Mont	25%
Pantolon, Çorap	50%
Pantolon, Mont	25%
Mont, Çorap	25%

Tablo 5. Destek Değeri Minimum Destek Değerinden Büyük 2-öğeli Kümeler

Öğeler	Destek Değerleri
Ayakkabı, Çorap	50%
Ayakkabı, Ceket	50%
Ayakkabı, Pantolon	50%
Ayakkabı, Mont	50%
Pantolon, Çorap	50%

Tablo 6’da 3-öğeli kümelerin destek değerleri verilmiştir. Tablo 6’daki 3-öğeli kümelerden minimum destek değeri 50%’den büyük olanlar ise Tablo 7’de verilmiştir.

Tablo 6. 3-öğeli Kümelerin Destek Değerleri

Öğeler	Destek Değerleri
Ayakkabı, Çorap, Ceket	25%
Ayakkabı, Çorap, Pantolon	50%
Ayakkabı, Çorap, Mont	25%
Ayakkabı, Ceket, Pantolon	25%
Ayakkabı, Ceket, Mont	25%
Ayakkabı, Pantolon, Mont	25%
Pantolon, Çorap, Mont	25%
Pantolon, Çorap, Ceket	25%

Tablo 7. Destek Değeri Minimum Destek Değerinden Büyük 3-öğeli Kümeler

Öğeler	Destek Değerleri
Ayakkabı, Çorap, Pantolon	50%

Minimum destek şartını sağlayan ürün (nesne) kümelerini bulduktan sonra birliktelik kurallarını oluşturabiliriz. Anlaşılır olması için sadece destek değeri minimum destek değerinden büyük 3-öğeli kümelerden kurallar oluşturalım. Kurallarımız, minimum güven değerinden daha büyük bir güven değerine sahip olmalıdır.

Tablo 8. 3-öğeli Sık Ürün (Nesne) Kümelerinden Oluşan Birliktelik Kuralları

Öğeler	Güven Değerleri
Ayakkabı, Çorap → Pantolon	100%
Ayakkabı, Pantolon → Çorap	100%
Pantolon, Çorap → Ayakkabı	100%

Ortaya çıkan değerlere göre ayakkabı ve çorap ürünlerini aynı anda alan kişi mutlaka (100%) pantolon ürününü de almaktadır. Aynı şekilde ayakkabı ve pantolon ürünlerini aynı anda alan kişi mutlaka çorap ürününü de almaktadır. Son olarak, pantolon ve çorap ürünlerini aynı anda alan kişi ayakkabıyı da almaktadır.

Elde ettiğimiz bu kurallar kullanılarak veri stok durumu ve raf düzen sistemi için düzenlemeler yaparak veri madenciliği tekniğini reel fayda getirecek bir biçimde kullanmış oluruz.

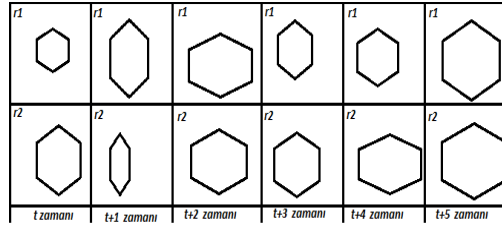
2.2. Hareketli Veri

Kablosuz iletişim araçlarının yaygın şekilde kullanımı, elde edilen zamansal-mekânsal verinin miktarını önemli ölçüde artırmaktadır (Verhein ve Chawla, 2008).

Nesneler uzayda sürekli olarak hareket halindedirler. Bu nesnelerin hareketlerini tespit etmek, kayıt altına almak ve sonraki hareketlerini tahmin etmek gibi çalışmalar günümüzde oldukça yaygındır. Ancak yapılan bu işlemler oldukça zor olmakla birlikte, veri tabanında kapladığı alan bakımından da “terabyte” hatta “petabyte” boyutuna ulaşmaktadır.

2.2.1. Nesne Hareketleri

Nesnenin farklı zamanlarda, farklı alan veya şekilde bulunması sonucu oluşan durumuna nesnenin hareketi denir. Hareketli nesneler zamanla değişen noktalar, çizgiler, alanlar veya hacimler olabilir (Macedo vd., 2008).

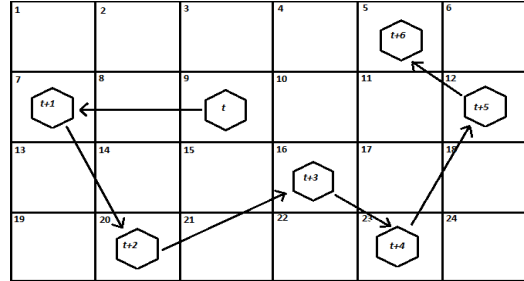


Şekil 2. Hareketli Nesnenin Zamanla Değişimi

Şekil 2’de $r1$ ve $r2$ alanlarındaki cisimlerin zamanla şekillerinde ortaya çıkan değişimler verilmektedir.

2.2.2. Hareketli Nesnenin Yörüngesi

Nesnenin yörüngesi, nesne her hareket ettikçe güncellenmektedir. Yani, iki boyutlu veya üç boyutlu uzayda nesnenin bulunduğu konum bilgilerinin değişmesi yörüngenin oluşması anlamına gelir. Konum bilgileri x-y iki boyutlu koordinat sistemi veya x-y-z üç boyutlu koordinat sistemiyle ifade edilebilir.



Şekil 3. Hareketli Nesnenin Yörüngesinin Zamanla Değişimi

Şekil 3'deki örnekte görüldüğü gibi cismin konumunun değişmesi sonucunda cismin yörüngesi zamanla değişmektedir. Klasik veri türleri ile cismi iki veya üç boyutlu olarak tanımlamak oldukça zor olmaktadır. Ancak veri tabanı yönetim sistemleri, mekânsal veriyi tanımlayabilmek için oluşturdukları yeni veri tipleri sayesinde verinin yörüngesini de tanımlayabilmişlerdir. Bu tanımlama genellikle verinin hareket yönünün sözel olarak ifade edilmesiyle oluşturulur (Örneğin; Rüzgarın esme yönü için: kuzey, güney, kuzeydoğu, güneybatı vb., Göl suyu hareketleri için: merkeze doğru, kıyılara doğru vb). Tanımlanan verinin yörüngesi sayesinde yörüngedeki değişimler incelenerek cismin yörünge eğilimi hakkında çıkarımlar yapılabilmektedir.

2.2.3. Hareketli Nesnenin Hareketinin Kaydedilmesi

Nesne hareketlerinin kaydedilmesi olayı, son 10 yılda çok önemli bir araştırma alanı haline gelmiştir (Macedo vd., 2008). Nesnelerin hareketlerinin kaydedildiği veritabanlarına yörüngesel veritabanı da diyebiliriz. Yörüngesel veri tabanları, hareketli nesnenin yörüngesinin sorgulanması ve sunulması desteği için veri tabanı teknolojisinin genişletilmesi ile ortaya çıkmıştır (Macedo vd., 2008). Mevcut veri tabanı yönetim sistemleri, bu genişlemeye imkân sundukları için, iki boyutlu veriyi veri tabanında modelleyerek hareketli nesnenin hareketini de tanımlanmışlardır. Böylece nesnenin hareket eğilimleri belirlenecek ve sonraki olası hareketleri tahmin edilebilecektir.

2.3. Zamansal-Mekânsal Veri

Zamansal-Mekânsal Veri Tabanı Yönetim Sistemleri (*Spatio-Temporal Database Management Systems, STDBMS*), konum (*spatial*) özellikleri zamana (*time*) bağılı olarak değişen nesnelerin modellenmesi, depolanması, erişimi ve sorgulanması fonksiyonlarını içermektedir. STDBMS, önemli uygulamalarda kullanılan bir teknolojidir. Günümüzde gerek ticari gerekse akademik düzeydeki bu uygulamalara örnek olarak Coğrafi Bilgi Sistemleri (*Geographic Information System, GIS*), çoklu ortam CAD uygulamaları, hareketli nesne takip ve sorgulama sistemleri verilebilir.

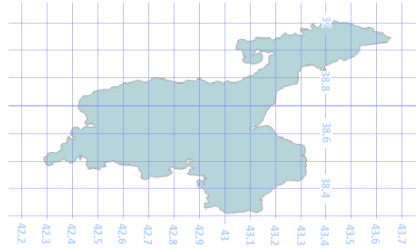
Zamansal-mekânsal veri, ya zamansal veritabanının ya da mekânsal veritabanının genişletilmesi ile modellenir. Zamansal veritabanına mekânsal özellikler, ya da mekânsal veritabanına zamansal özellikler ekleyebiliriz. Bu çalışmada, zamansal veritabanına mekânsal özellikleri ekleme yöntemi kullanılmaktadır. Böylece zaman içerisinde mekânda meydana gelen değişiklikleri analiz ederek, veriyi zaman-mekân boyutunda inceleme imkanı bulabilmekteyiz.

2.3.1. Zamansal Veri

Verinin tarih ve saat bilgisi olarak veri tabanında tutulması sonucu oluşmuş halidir. Böylece nesnemizin zamansal olarak sorgulanması ve sunulması mümkün olacaktır.

2.3.2. Mekânsal Veri

Mekânsal bir öğenin koordinatlarının mekânsal bir veri türünde (*geography, geometry vb.*) veri tabanında tutulması sonucu oluşmuş halidir. Böylece nesnemizin mekânsal olarak sorgulanması ve sunulması mümkün olacaktır.



Şekil 4. Van Gölü'nün Konum Bilgileri

Şekil 4’de örnek olarak Van Gölü’nün kıyı koordinatlarının yaklaşık 3000 kıvrım noktasının belirtilmesiyle gölün konumsal (mekânsal) verisi ifade edilmiştir. Elimizdeki konumsal veriler uygun bir şekilde veri tabanına kaydedilerek gölün haritası belirtilmiş olur.

2.3.3. Zamansal-Mekânsal Veri Tabanı Yönetim Sistemleri

Verilerin daha hızlı ve tutarlı olarak işlenmesi için genellikle büyük yazılım firmaları tarafından geliştirilmiş sistemlerdir. Bu sistemler sayesinde çok kullanıcı yapılar desteklenmiş ve geliştirilmiştir. IBM-DB2, Informix, Oracle-Oracle Spatial ve Microsoft-SQL Server gibi veri tabanı yönetim sistemleri mekânsal veri tanımlama ihtiyacı duymuşlardır ve “*geography*”, “*geometry*” vb. veri türlerini sistemlerine eklemişlerdir.

2.3.4. Zamansal-Mekânsal Verinin Kaydedilmesi

Zamansal-Mekânsal Veri Tabanı Yönetim Sistemleri (ZMVTYS) sayesinde mekânsal veriler kaydedilmekte ve sorgulanabilmektedir. Ayrıca zaman boyutunun (özellikinin, alanının) da eklenmesiyle üç boyutlu veri işlenmesi mümkün hale gelmiştir.

Aşağıdaki örneğimizde “vanGolu” adında bir tablo oluşturularak, alanları ve veri türleri tanımlanmaktadır. Geleneksel veri türlerindeki alanlara ek olarak “*geography*” veri türünde “konum” adında bir alan tanımlanması da yapılmıştır. Bu durum, mekânsal bir veri türünün (burada *geography*) kullanımına örnek teşkil etmektedir. Ayrıca “tarih” adında, zaman veri türünde bir alanın tanımlanması ise veritabanı tablomuza zamansal bir boyut katmaktadır. Tablomuzun diğer alanları (id, şehir, alan, çevre, hareketYonu) ise standart veri türlerine sahiptir.

CREATE TABLE [vanGolu]

([id] [int] NOT NULL, [şehir] [varchar](50) NULL, [konum] [*geography*] NULL, [alan] [varchar](50) NULL, [çevre] [varchar](50) NULL, [hareketYonu] [varchar](50) NULL, [tarih] [Date])

Tablomuzu tanımladıktan sonra konumsal verileri aşağıda gösterildiği üzere kolaylıkla kayıt altına alabiliriz.

INSERT INTO [vanGolu]

([id], [sehir], [konum], [alan], [cevre], [hareketYonu], [tarih])

VALUES

(1, 'Van', geography::STGeomFromText('MULTIPOLYGON

((38.4240080236828 42.91397094726564, 38.91240739487224
43.58413696289062,

.

38.91507876347639 43.57589721679687, 38.9156130251313
43.57246398925781,

38.4240080236828 42.91397094726564))', 4326), '3626000000', '450000',
'disYoneTaraf', '2009-07-20');

2.3.5. Zamansal-Mekânsal Verinin Sorgulanması

Belirli bir mekâna ait belirleyici özellikler zaman içinde değişim gösterebilir. Bahsi geçen mekân için, herhangi bir zamandaki bilgilere ulaşmak için uygun SQL komutlarını yazabiliriz. Örneğin:

SELECT sehir, konum, alan, cevre, hareketYonu, tarih **FROM** vanGolu
where tarih = '2009-07-20'

2.4 Apriori Algoritmasının Zamansal-Mekânsal Verilere Uyarlanmasına Örnek

Apriori algoritmasını, elimizde konum bilgileri ve göl suyu seviyesini etkileyen değişken bilgileri olan zamansal-mekânsal veri karakterine sahip Van Gölü örneğine uyarlayalım. Amacımız, göl suyu seviyesinin değişimini inceleyerek, su seviyesinin azalıp artmasına bağlı olarak kıyı kesimlerdeki taşmaları veya sığ alanların kuruyarak yeni toprak arazileri oluşumuna neden olmalarını tespit etmektir.

Çalışmamızda, Van Gölü su seviyesini etkileyen 6 öge (değişken) kullanılmıştır: yağış, buharlaşma, yer üstünden göle akan sular, yer altından göle akan sular, yer altına gölden akan sular ve göl suyu kullanımı. Burada değişkenleri miktarlarına göre, basit anlamda "var-yok" şeklinde ifade edeceğiz ve bu değişkenler yardımıyla göl seviyesinin değişmesini inceleyebiliriz.

Tablo 9'da göl suyu seviyesini etkileyen öğelerin belli bir tarih aralığında meydana gelip gelmedikleri verilmiştir. Apriori

Algoritması'nın minimum destek değerini 30%, minimum güven değerini de 40% alarak, bu birliktelik kuralı öğrenme algoritmasını çalışmamıza uyarlayabiliriz:

2.1.3.3.1 kısmında anlatılan Apriori algoritmasının işleyişi aynen uygulanarak sırasıyla destek değeri minimum destek değerinden büyük 1-öğeli kümeler, 2-öğeli kümeler ve son olarak 3-öğeli kümeleri elde ederiz.

Tablo 9. Van Gölü'ne Ait Göl Seviyesini Etkileyen Değişkenlerin Gün Bazında Meydana Gelip Gelmemesi Durumu

İşlem ID(Tarih Bilgisi)	Gözlemlenen Değişkenler
2001-07-20	yağış, buharlaşma, yer üstünden göle akan sular
2001-07-21	buharlaşma, yer üstünden göle akan sular
2001-07-22	yer altına gölden akan sular, göl suyu kullanımı
...	...
2012-03-30	yağış, buharlaşma, yer üstünden göle akan sular, yer altından göle akan sular

Tablo 10'da göl suyu seviyesini etkileyen öğelerin her birinin tek-tek ele alınmasıyla bulunan destek değerleri verilmiştir. Bunlara, 1-öğeli küme destek değeri adı verilir. Örneğimize göre, bütün 1-öğeli kümeler minimum destek şartını sağlamaktadırlar.

Tablo 10. 1-öğeli Kümelerin Destek Değerleri

Öğeler	Destek
yağış	80%
buharlaşma	85%
yer üstünden göle akan sular	80%
yer altından göle akan sular	70%
yer altına gölden akan sular	65%
göl suyu kullanımı	50%

Tablo 11'de ise minimum destek şartını sağlayan 3-öğeli kümeler verilmiştir.

Tablo 11. Destek Değeri Minimum Destek Değerinden Büyük 3-öğeli Kümeler

Öğeler	Destek
yağış, yer üstünden göle akan sular, buharlaşma	50%
yağış, yer altından göle akan sular, buharlaşma	55%
buharlaşma, yer üstünden göle akan sular, yer altından göle akan sular	45%

Minimum destek (30%) şartını sağlayan nesne kümelerini bulduktan sonra birliktelik kurallarını oluşturabiliriz. Anlaşılır olması için sadece destek değeri minimum destek değerinden büyük 3-öğeli kümelerden kurallar oluşturalım. Kurallarımız, minimum güven değerinden (40%) daha büyük bir güven değerine sahip olmalıdır.

Tablo 12. Güven Değeri Minimum Güven Değerinden Büyük Bazı Birliktelik Kuralları

Birliktelik	Güven
yağış, yer üstünden göle akan sular → buharlaşma (göl seviyesi azalır)	100%
buharlaşma, yağış → yer altından göle akan sular (göl seviyesi artar)	50%
buharlaşma, yer üstünden göle akan sular, → yer altından göle akan sular (göl seviyesi artar)	50%

Tablo 12’de 3-öğeli kümelerden oluşan, minimum güven değerinden daha büyük güven değerine sahip bazı birliktelik kuralları verilmiştir.

Birliktelik kurallarının sonuç (sağ) tarafını inceleyerek su seviyesinin azalıp, azalmayacağını anlayabiliriz. Su seviyesinin artmasına bağlı olarak kıyı kesimlerde taşmalar, azalmasına bağlı olarak ise sığ alanların kuruyarak yeni toprak arazileri oluşturmaları muhtemeldir.

Sonuç olarak, elde edilen birliktelik kurallarını kullanarak, zaman içinde göl seviyesini etkileyen faktörlerin hangi tür etkileşimlerinin göl seviyesini artırıp artırmayacağını tahmin edebiliriz.

3. BULGULAR

Anlatılan yöntem ve metodlar kullanılarak nesnelerin zamansal ve mekânsal değişimleri incelenmiştir. Bu incelemeler sonucunda daha sonra meydana gelebilecek durumları tahmin edebiliriz. Özellikle, birliktelik kuralları öğrenme yöntemiyle açık bir şekilde görünmeyen ilişkileri ortaya koyarak muhtemel riskleri önleyebilir veya muhtemel yatırım alanlarını tespit edebiliriz.

4. SONUÇLAR VE TARTIŞMA

Günümüzde hemen her alanda veriler kayıt altına alınmaktadır. Kayıt altına alınan bu veriler zamanla oldukça büyük alanlar kaplamakta ve anlam ifade etmeyen bir yığın oluşturmaktadır. Bu anlamsız veri yığınının faydalı bir hale dönüştürülmesi için teknolojik gelişmeler bize imkân sunmaktadır. Özellikle yazılım alanında meydana gelen gelişmeler ile geçmişte kayıt altına alınmış verilerden çok önemli çıkarımlar yapmak mümkündür. Günümüzün Microsoft, Google, Nasa, Amazon gibi teknoloji devlerinin başarılarının altında veri madenciliği ve benzeri disiplinler yatmaktadır. Birliktelik kuralları öğrenme tekniği ise veri madenciliğinin önemli tekniklerinden birisi olarak karşımıza çıkmaktadır.

Apriori algoritması, birliktelik kuralları öğrenme tekniğinde oldukça yaygın olarak kullanılmaktadır. Klasik veri madenciliği alanında yaygın kullanıma sahip olan Apriori algoritması, zaman ve mekâna bağlı olarak değişen veri kümelerinde çok fazla kullanılmamaktadır. Zaman ve mekâna bağlı veri kümelerinin klasik veri kümelerinden farkı, verilerin klasik veri özelliklerine zaman ve konum özelliklerinin de eklenmesidir. Bu çalışmada Apriori algoritması, zamansal-mekânsal veri madenciliği alanına uyarlanarak, Van Gölü'ne ait zamansal-mekânsal veri kümesi üzerinde uygulanmıştır. Uygulama C# programlama dili ve SQL Server 2008 R2 veritabanı programı kullanılarak geliştirilmiştir. C# programlama dilinin mekânsal veriyi destekleyen kütüphanelerinin olması ve SQL Server 2008 R2 veritabanı programının ise mekânsal veri tipi desteğinden dolayı çalışmamızda kullanılmaları uygun bulunmuştur.

Çalışmamızda, Van Gölü su seviyesini etkileyen 6 öge kullanılmıştır: yağış, buharlaşma, yer üstünden göle akan sular, yer altından göle akan sular, yer altına gölden akan sular ve göl suyu kullanımı. Burada değişkenler miktarlarına göre, basit anlamda “var-yok” şeklinde ifade edilmiştir ve bu değişkenler yardımıyla göl seviyesinin değişmesi incelenmiştir.

Göl suyu seviyesini etkileyen öğelerin belli bir tarih aralığında meydana gelip gelmedikleri verileri kullanarak, Apriori Algoritması'nın minimum destek değeri 30%, minimum güven değeri de 40% alarak, birliktelik kuralı öğrenme algoritması söz konusu verilerimize uygulanmıştır.

Birliktelik kurallarının sonuç (sağ) tarafını inceleyerek su seviyesinin azalıp, azalmayacağını anlayabiliriz. Su seviyesinin artmasına bağlı olarak kıyı kesimlerde taşmalar, azalmasına bağlı olarak ise sığ alanların kuruyarak yeni toprak arazileri oluşturmaları muhtemeldir.

Sonuç olarak, elde edilen birliktelik kurallarını kullanarak, zaman içinde göl seviyesini etkileyen faktörlerin hangi tür etkileşimlerinin göl seviyesini artırıp artırmayacağını tahmin edebiliriz.

5. ÖNERİLER

Gelişen dünya ülkeleri ile rekabet etmek için özellikle teknolojiye geri kalmamak gereklidir. Bu nedenle değişen ve gelişen teknolojiyi sürekli olarak takip etmek gereklidir. Ülke olarak diğer ülkelerle paralel bir yaşam tarzına sahip olmak için teknoloji alt yapısını daha da iyileştirmeli ve insanları bu teknolojileri kullanmaya ve geliştirmeye teşvik etmeliyiz.

Veri madenciliği, gelişmiş ülkelerde birçok sektörde verimli bir biçimde kullanılmakta olup, ülkemizde de yavaş yavaş kullanımı artan bir bilim dalıdır. Öğrenme ve akabinde de tahminde bulunma süreçlerini içeren her türlü gerçek dünya problemlerinde veri madenciliği teknikleri kullanılmalıdır.

TEŞEKKÜR

Coğrafi bilgi sistemiyle alakalı olarak kaynak temin eden Yrd. Doç. Dr. Faruk ALAEDDİNOĞLU'na teşekkür ederiz.

KAYNAKLAR

- Agrawal, R., Imielinski, T. ve Swami, A., (1993). Mining Association Rules Between Sets of Items in Large Databases, *In Proceedings of the ACM SIGMOD International Conference on Management of Data*, 207-216, Washington, USA.
- Agrawal, R. ve Srikant, R., (1994). Fast Algorithms for Mining Association Rules, *In Proceedings of the 20th International Conference on Very Large Databases*, 487-489, Santiago, Chile.
- Agrawal R. ve Srikant R., (1995). Mining Sequential Patterns, *11th International Conference on Data Engineering*, 3-14, Taipei, Taiwan.
- Das A., Ng, W. K. ve Woon Y. K., (2001). *Rapid Association Rule Mining*, In *Proceedings of the Tenth International Conference on Information and Knowledge Management*, ACM Press, 487-499, Atlanta, GA, USA.
- Han, J. ve Kamber, M., (2006). *Data Mining Concepts and Techniques*, (s.1-35) San Francisco, Morgan Kauffmann Publishers Inc.,.
- Houtsma, M. ve Swami, A, (1995), Set-Oriented Mining for Association Rules in Relational Databases, *Proceedings of the 11th IEEE International Conference on Data Engineering*, 25-34, Taipei, Taiwan.
- Karabatak M. & İnce M.C., Apriori Algoritması ile Öğrenci Başarı Analizi, *Emo Bilimsel Dergi*, 1, 12.01.2012 tarihinde http://www.emo.org.tr/ekler/24f4c5eef7ec01c_ek.pdf adresinden alınmıştır.
- Macedo J., Vangenot C., Othman W., Pelekis N., Frentzos E., Kuijpers B., Ntoutsis I., Spaccapietra S. ve Theodoridis Y., 2008. *Trajectory Data Models*, Springer
- Olson, D.L. ve Delen D., (2008). *Advanced Data Mining Techniques*, 1. Baskı, 7. S., German, Springer
- Özçakır F.C. ve Çamurcu A.Y., 2007. Birliktelik Kuralı İçin Bir Veri madenciliği Yazılım Tasarımı ve Uygulaması, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Sayı:12, s. 21-37
- Özkan Y., (2008). *Veri madenciliği Yöntemleri*, 3, 1. Baskı, 3, 40, İstanbul, Papatya Yayıncılık
- Sever, H. ve Oğuz, B., 2002. Veritabanlarında Bilgi Keşfine Formal Bir Yaklaşım, Kısım 1: Eşleştirme Sorguları ve Algoritmalar, *Bilgi Dünyası*, 3, 2, 173-204.

- Sack R.D.,2010, *Geography, Geometry and Explanation*, Routledge, 1, 14.03.2012 tarihinde <http://dx.doi.org/10.1111/j.1467-8306.1972.tb00844.x> adresinden alınmıştır.
- Şentürk A., (2006). *Veri madenciliği: Kavram ve Teknikler*, 2, 1. Baskı, 19, Bursa, Ekin Basın Yayın
- Verhein F. ve Chawla S., 2008. *Mining Spatio-Temporal Patterns In Object Mobility*, Springer
- Zaki, M. J. ve Hsiao, C. J., (2002). CHARM: An Efficient Algorithm for Closed Itemset Mining, *In 2nd SIAM International Conference on Data Mining* , Eds. Grossman, R. L., Han, J., Kumar, V., Mannila, H. ve Motwani, R., s.457-473,Siam, Arlington, VA, USA,.
- Zhu, H., (1998). *On-Line Analytical Mining of Association Rules*, MSc. Thesis, Simon Fraser University, Ottawa, Canada.
