MEKÂNSAL UYGULAMALAR İÇİN VERİ MADENCİLİĞİ YAKLAŞIMI

Mevlüt UYAN[1], Tayfun ÇAY[2]

[1]Selçuk Üniversitesi, Yapı İşleri ve Teknik Daire Bşk., muyan@selcuk.edu.tr
[2]Selçuk Üniversitesi, Jeodezi ve Fotogrametri Müh. Bölümü, tcay@selcuk.edu.tr

ÖZET

Teknolojik gelişmelere bağlı olarak bilgisayar sistemlerinin her geçen gün güçlerini artırması ile bilgisayar sistemlerinde sayısal bilgi miktarı da aynı şekilde artmıştır. Bu gelişmeye paralel olarak veri tabanları da kapasitelerini arttırmıştır ve daha fazla veri veritabanlarında saklanmaktadır. Bu sayede doğru ve daha detaylı bilgiye ulaşmamız mümkün hale gelmiş fakat ortaya çıkan bu büyük sayısal veri yığınlarının yönetilmesi ve anlamlı hale getirilmesi sorunu ortaya çıkarmıştır. Bu verilerin istenilen amaç doğrultusunda bilgiye veya anlamlı hale dönüştürme işlemi veri madenciliği metotları ile yapılabilmektedir.

Veritabanlarındaki tüm verinin % 80'inin mekânsal bileşenlere sahip olduğu düşünülürse mekânsal verilerden bilgi keşfi içinde veri madenciliğinin önemi daha iyi anlaşılabilir. Bu konuda yapılan çalışmalar Mekânsal veri Madenciliği adı altında son yıllarda büyük bir gelişme kaydetmiştir. Çalışmada veri madenciliği ve mekânsal veri madenciliğinde kullanılan temel kavramlardan bahsedilmiştir.

Anahtar Sözcükler: Veri Madenciliği, Mekânsal Veri Madenciliği.

ABSTRACT

DATA MINING APPROACH FOR SPATIAL APPLICATIONS

As more and more computer systmens have increased their power due to technological advances, amount of numerical knowledge in computer systems has increased as well. In paralel to this advance, databases have increased their capacities, and much more data has been kept in databases. Thus, We have become able to access correct and more detailed information, but the problems consorning the control and make meaning of these huge numerical mass data have been emerged. The operation which is turning them into making meaning or knowledge in terms of aim can be made by data mining methods.

Given that 80% of all data has spatial combination, the important of data mining can be understood in discovering knowledge of spatial data. Recently the study made at this subject has great developed under the name of spatial data mining. In this study fundamental topics used in spatial data mining and data mining have been mentioned.

Keywords: Data Mining, Spatial Data Mining.

1. GİRİŞ

Otomatik veri toplama araçları ve veri tabanı teknolojilerindeki gelişme, veritabanlarında ve diğer bilgi depolarında çok miktarda bilgi depolanması sonucunu doğurmuş tam anlamıyla bir veri patlaması yaşanmıştır. Verilerin dijital ortamda saklanmaya başlanması ile birlikte, yeryüzündeki bilgi miktarının her 20 ayda bir kendini iki katına çıkardığı günümüzde, veri tabanlarının sayısı da benzer, hatta daha yüksek oranda artmaktadır (Vahaplar, İnceoğlu, 2001). Yüksek kapasiteli işlem yapabilme gücünün ucuzlamasının bir sonucu olarak, veri saklama hem daha kolay olmuş, hem de verinin kendisi de ucuzlamıştır.

Varolan verilerden bilgiyi yani kullanılabilecek örüntüleri elde etmeye geniş çapta ihtiyaç duyulmuştur. Bu ihtiyacı gidermek amacıyla araştırma kurumları ve üniversiteler çalışmalarıyla yeni disiplinler ortaya çıkarmıştır. Veri madenciliği bu yeni disiplinlerden biridir. Veri madenciliğinin veri tabanları üzerine uygulanmasıyla Veri Tabanlarında Bilgi Keşfi (VTBK) ortaya çıkmıştır (Aydoğan, 2003).

Veri madenciliğini istatistiksel bir yöntemler serisi olarak görmek mümkündür. Ancak veri madenciliği, geleneksel istatistikten birkaç yönde farklılık gösterir. Veri madenciliğinde amaç, kolaylıkla mantıksal kurallara ya da görsel sunumlara çevrilebilecek nitel modellerin çıkarılmasıdır. Veri madenciliği sahası, istatistik, yapay

zeka, veri tabanları ve veri görselleştirme gibi alanlar ile yakından ilişkilidir. Veri madenciliği ve bilgi keşfi, özellikle elektronik ticaret, bilim, tıp, iş ve eğitim alanlarındaki uygulamalarda yeni ve temel bir araştırma sahası olarak ortaya çıkmaya başlamıştır (Baykasoğlu, 2005).

Gelişen teknoloji ile birlikte Coğrafi Bilgi Sistemlerinden (CBS) hızlı veri toplama işlemi insanların yaptığı analiz kapasitesini aşan ve çok sayıda coğrafi katman içeren yüksek kapasiteli mekansal veri tabanlarının ortaya çıkmasına sebep olmuştur (Castro, Lee, 2004). Bu nedenle son çalışmalar Mekansal Veri Madenciliği altında CBS'nin keşifsel ve analitik kapasitesini arttırmak için otomatik ve yarı otomatik araçlar geliştirmeye yönelmiştir (Openshaw, 1999).

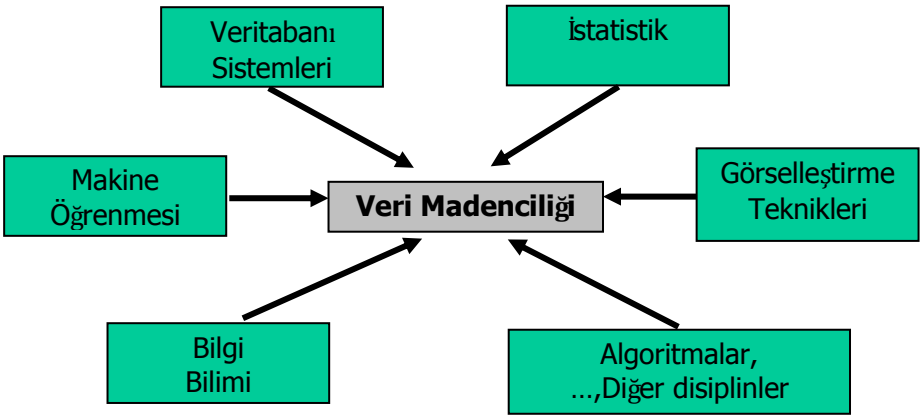
Bu çalışmada veri madenciliği ve mekansal veri madenciliği ile ilgili temel kavramlardan bahsedilmiştir.

2. VERİ MADENCİLİĞİ

Son zamanlarda kurumlar çok büyük miktarda bilgiyi veritabanlarında toplamış ve burada gizli bulunan bilginin potansiyelini fark etmeye başlamışlardır. Çok büyük miktarda veri içeren bu veritabanlarından, ilginç (sıra dışı, önemli, kesin, önceden bilinmeyen ve yararlı olabilecek) bilgilerin ya da örüntülerin bulunup çıkarılması işlemi veri madenciliği yaklaşımının özüdür. Pek çok kaynakta veri madenciliği "*Veritabanlarında bilgi bulma*" *(KDD-Knowledge Discovery in Databases)* olarak geçmektedir (Ester ve diğ., 1995, Wang ve diğ., 2005, Miller, H. J., Han, J., 1999).

Temel olarak veri madenciliği bilgisayar destekli bir bilgi çözümleme işlemidir. Veri madenciliği, eldeki verilerden çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasıdır. Bu işlem kümeleme, veri özetleme, değişikliklerin analizi, sapmaların tespiti gibi belirli sayıda teknik yaklaşımları içerir. Başka bir deyişle, veri madenciliği, verilerin içerisindeki örüntülerin, ilişkilerin, değişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların keşfedilmesidir (Vahaplar, İnceoğlu, 2001).

Veri madenciliği, veri setleri arasındaki örüntülerin ya da düzenin, verinin analizi ve yazılım tekniklerinin kullanılması ile ilgilidir. Veriler arasındaki ilişkiyi, kuralları ve özellikleri belirlemekten bilgisayar sorumludur. Amaç, daha önceden fark edilmemiş veri desenlerini tespit edebilmektir. Bir başka tanıma göre ise, istatistik ve matematik tekniklerle birlikte örüntü tanıma teknolojilerini kullanarak depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni korelasyon, örüntü ve eğilimlerin keşfedilmesi sürecidir. Veri madenciliğini istatistiksel bir yöntemler serisi olarak görmek mümkün olabilir. Ancak veri madenciliği, geleneksel istatistikten birkaç yönde farklılık gösterir. Veri madenciliğinde amaç, kolaylıkla mantıksal kurallara ya da görsel sunumlara çevrilebilecek modellerin çıkarılmasıdır. Bu bağlamda, veri madenciliği insan merkezlidir ve bazen insan-bilgisayar ara yüzü birleştirilir (Vahaplar, İnceoğlu, 2001).



Şekil 1: Veri madenciliğinin, birçok disiplinle çakışma noktaları vardır (Yubin, 2005).

2.1 Veri Madenciliği Süreci

Bu süreç genelde beş temel adımı kapsar (Wu, Lu, 2002).

- ***Seçim:*** Çeşitli kaynaklardan veri elde etme.
- ***Önişleme:*** Temizlenmiş veri.
- ***Dönüşüm:*** Ortak formata dönüşüm. Yeni formata dönüştürme
- ***Veri madenciliği:*** İstenen sonuçları elde etme.
- ***Yorum/Değerlendirme:*** Anlaşılır biçimde kullanıcılara sonuçların sunulması.

Veri Seçimi aşaması verinin tanımlanarak çıkarılmasını kapsar. Bundan sonra uygun girdi özellikleri seçilir ve çıktı bilgileri sunulur. büyük veri ambarlarından verinin çıkarılması, işlenmesi ve analizi için araçların etkin olarak kullanılması ve uygulanması önemli uzmanlık gerektirir.

Veri Filtreleme ve Ön İşleme, kayıp veri alanlarının çıkarılması için stratejiler, model için gerekli bilgileri toplama, bilinen değişiklikleri hesaplama, verileri normalleştirme işlemleridir.

Veri Dönüşümü işlemi, istenen yollardan verilerin düzenlenmesini, verinin bir formattan farklı bir formata dönüştürülmesini, yeni özelliklerin tanımlanmasını ve veri boyutunun küçültülmesini kapsar.

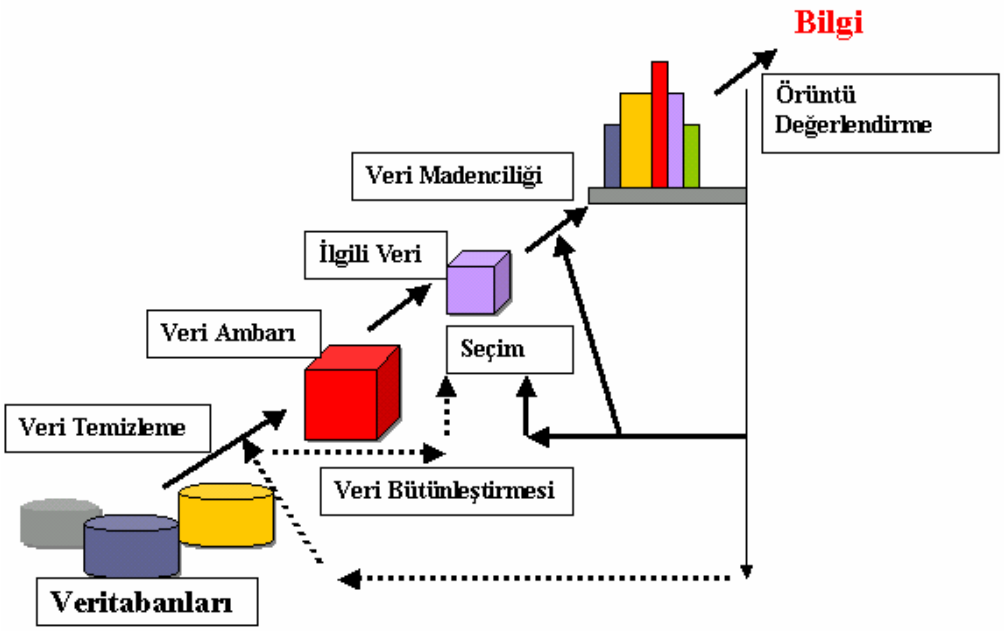
Bilgi Keşfi aşaması bir veya daha fazla teknik kullanılarak ilginç modellerin ortaya çıkarılması işlemini kapsar.

Yorum ve Onaylama aşaması sentezlenen bilginin verdiği mesajın anlaşılması işlemidir. Çıkarılan bilgiler uygulama alanındaki ilk uzmanlar tarafından karşılaştırma yoluyla değerlendirilir.

2.2. Veri Madenciliğinde Karşılaşılan Problemler

Veri madenciliği girdi olarak ham veriyi sağlamak üzere veri tabanlarına dayalı bir sistem olduğu için bu durum veri tabanlarının dinamik, eksiksiz, geniş ve net veri içermemesi durumunda sorunlar doğurur. Diğer sorunlar da verinin konu ile uyumsuzluğundan doğabilir. Sınıflandırmak gerekirse başlıca sorunlar şunlardır (Vahaplar, İnceoğlu, 2001).

Sınırlı Bilgi **:** Veri tabanları genel olarak veri madenciliği dışındaki amaçlar için tasarlanmışlardır. Bu yüzden, öğrenme görevini kolaylaştıracak bazı özellikleri içermeyebilir.



Şekil 2: Veri Madenciliği Süreci (Baykal, 2003).

Gürültü ve Eksik Değerler **:** Veri özellikleri ya da sınıflarındaki hatalara gürültü denir. Veri tabanlarındaki eksik bilgi ve bu yanlışlardan dolayı veri madenciliği amacına tam olarak ulaşmayabilir. Bu bilgi yanlışlığı, ölçüm hatalarından, ya da öznel yaklaşımdan olabilir.

Belirsizlik: Yanlışlıkların şiddeti ve verideki gürültünün derecesi ile ilgilidir. Veri tahmini bir keşif sisteminde önemli bir husustur.

Ebat, güncellemeler ve konu dışı sahalar **:** Veri tabanlarındaki bilgiler, veri eklendikçe ya da silindikçe değişebilir. Veri madenciliği perspektifinden bakıldığında, kuralların hala aynı kalıp kalmadığı ve istikrarlılığı problemi ortaya çıkar.

2.3 Veri Madenciliğinde Kullanılan Teknikler

Veri madenciliği sürecinde ilk yapılması gereken şey bilgi ihtiyacının belirlenmesi ve problemin tanımlanmasıdır. Bu aşamadan sonra veri kaynaklarının belirlenmesi ve verilerin incelenmesi gerekir. Veriler hazırlanır. Bundan sonraki aşama modelin oluşturulmasıdır. İşlemin en önemli adımı sayılabilecek bu aşama gerekli bilgiye ulaşmaya giden yoldur. Bundan sonrası modelin değerlendirilmesi ve sonuçların yorumlanması adımıdır.

Veri madenciliğinde kullanılan modeller, tanımlayıcı *(Descriptive) ve* tahmin edici *(Predictive)* olarak iki ana başlık altında incelenmektedir. Tanımlayıcı modellerde, karar vermeye yardımcı olarak kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. Tahmin edici modellerde, sonuçları bilinen verilerden hareketle bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır (Eker, 2007).

Gerek tanımlayıcı gerekse tahmin edici modellerde yoğun olarak kullanılan belli başlı isatistiki yöntemler; Sınıflama (Classification) ve Regresyon (Regression), Kümeleme (Clustering), Birliktelik Kuralları (Association Rules), bellek tabanlı yöntemler, yapay sinir ağları ve karar ağaçları olmak üzere altı ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir (Eker, 2007).

2.3.1 Sınıflama (Classification) ve Regresyon (Regression)

Sınıflama ve regresyon, önemli veri sınıflarını ortaya koyan veya gelecek veri eğilimlerini tahmin eden modelleri kurabilen veri analiz yöntemleridir. Sınıflama kategorik değerleri tahmin ederken, regresyon mevcut değerleri kullanarak diğer değerlerin ne olacağını tahmin etmeye çalışır.

Amaç, yeni bir nesnenin, belirli sınıflar içinde hangi sınıfa ait olup olmadığını belirleyecek bir sınıflayıcı (classifier) oluşturmaktır. Sınıflayıcı, gerileme modellerinin özel bir biçimi olarak kabul edilebilir (Baykal, 2003).

Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler şunlardır (Eker, 2007):

- Genetik Algoritmalar (Genetic Algorithms),
- K-En Yakın Komşu (K-Nearest Neighbor),
- Naive-Bayes,
- Çoklu Regresyon, Lojistik Regresyondur (Logistic Regression),
- Faktör ve Ayırma analizleri (factor and discriminant)

2.3.2 Kümeleme (Clustering)

Temel amaç, çok boyutlu veriler içindeki doğal grupları (kümeleri) bulmaktır. Nesneler, birbirlerine benziyorlarsa (aynı ölçüye göre) ve başka kümelerdeki nesnelere benzemiyorlarsa, aynı kümeye alınabilir. Kümelemede, alan bilgisinin kümeleme mekanizmalarıyla nasıl birleştirilebileceğiyle ilgilenilir (Baykal, 2003). Bazı uygulamalarda kümeleme modeli, sınıflama modelinin bir önişlemi gibi görev alabilmektedir.

Veri tabanlarında toplanan veri miktarındaki artışla, kümeleme analizi son zamanlarda veri madenciliği araştırmalarında aktif bir konu haline gelmiştir.

Kullanılacak olan kümeleme algoritmasının seçimi, veri tipine ve amaca bağlıdır. Genel olarak başlıca kümeleme yöntemleri şu şekilde sınıflandırılabilir:

- Bölme yöntemleri (Partitioning methods)
- Hiyerarşik yöntemler (Hierarchical methods)
- Yoğunluk tabanlı yöntemler (Density-based methods)
- Izgara tabanlı yöntemler (Grid-based methods)
- Model tabanlı yöntemler (Model-based methods)

2.2.3 Birliktelik Kuralları (Association Rules)

Amaç birlikte gerçekleşmesi muhtemel olayların kurallarının ortaya konmasıdır. Birliktelik kuralları, veri kümeleri arasında birliktelik ilişkileri bulurlar. Büyük miktardaki mesleki işlem kayıtlarından ilginç birliktelik ilişkilerini keşfetmek, şirketlerin karar alma işlemlerini daha verimli hale getirmektedir. Büyük veri tabanlarında birliktelik kuralları bulunurken öncelikle sık tekrarlanan öğeler bulunur. Bu öğelerin her biri en az, önceden belirlenen minimum destek sayısı kadar sık tekrarlanırlar. Daha sonra sık tekrarlanan öğelerden güçlü birliktelik kuralları oluşturulur. Bu kurallar minimum destek ve minimum güven değerlerini karşılamalıdır.

2.3.4 Bellek Tabanlı Yöntemler

Bellek tabanlı yöntemler istatistikte 1950'li yıllarda önerilmiş olmasına rağmen o yıllarda gerektirdiği hesaplama ve bellek yüzünden kullanılamamış ama günümüzde bilgisayarların ucuzlaması ve kapasitelerinin artmasıyla, özellikle kullanılabilir olmuştur. Bu yönteme en iyi örnek en yakın k komşu algoritmasıdır (k-nearest neighbor) (Eker, 2007).

2.3.5 Yapay Sinir Ağları(YSA)

Bu yöntemler veri madenciliğinin en etkili yöntemleri arasında yer alan yapay sinir ağları, genetik algoritmalar, genetik programlama, bulanık mantık gibi teknikleri içerir. Yapılan pek çok çalışma bu tekniklerin özellikle doğrusal olmayan, bulanık ve değişken yapıdaki veri uzaylarında klasik istatistik temelli yöntemlere göre çok daha başarılı sonuçlar verdiğini göstermektedir (Baykasoğlu, 2005).

2.3.6 Karar Ağaçları

Karar ağaçları, veri madenciliğinde kuruluşlarının ucuz olması, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca entegre edilebilmeleri nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahip tekniktir. Karar ağacı, adından da anlaşılacağı gibi bir ağaç görünümünde, tahmin edici bir tekniktir. Karar ağaçları kural çıkarma algoritmalarıdır. Bu algoritmalar bir veri kümesinden kullanıcıların çok kolay anlayabileceği (IF-THEN) türündeki kuralları bir ağaç yapısında türetebiliriler (Baykasoğlu, 2005).

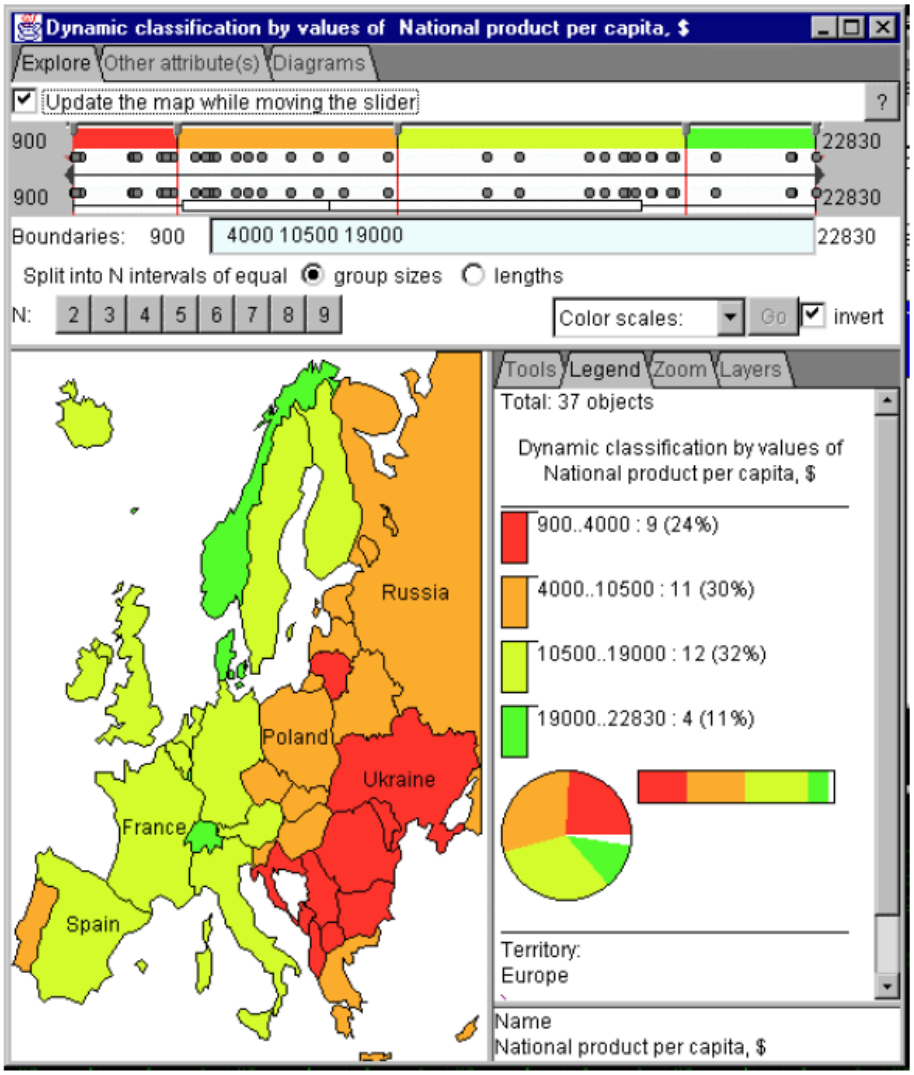
3. MEKÂNSAL VERİ MADENCİLİĞİ

Mekansal Veri Madenciliği (MVM) kısaca veri madenciliği tekniklerinin mekansal verilere uygulanmasıdır. Veri madenciliği büyük veri tabanlarında var olabilen gizli modelleri arayıp bulur. MVM de mekansal veri tabanlarında var olabilen gizli özellikler ve ilginç ilişkileri keşfeder. Mekansal tahmin yükleme için uzman bölgelerin bulunması ve mekansal unsurların kesin etkilerin tanımlanması arazinin yeniden geliştirilebilmesi için uzun süreli gözlemler yapılması gerektiğinden zordur (Wu, Lu, 2002).

Veri toplama ve işleme teknolojilerinin gelişmesinden dolayı son yıllarda sayısal coğrafik veri setlerinin miktarı ve kapsamı da çok hızlı büyümüştür. Bu durum yüksek çözünürlüklü uzaktan algılama sistemleri ve diğer izleme aygıtları ile coğrafi olarak referanslandırılmış sayısal görüntülerin çok miktarda elde edilmesini de kapsamaktadır. Mekansal uygulamalar çok büyük hacimli veriler üretmektedir. Bu sebepten dolayı, mekansal ilişkilerin hesaplanması çok zaman almaktadır. Dijital verinin % 80'i mekansal bileşenlere sahiptir (Shekhar, Chawla, 2003). Geleneksel mekânsal analiz metotları etkin olarak sadece sınırlandırılmış ve aynı türden verileri işleyebilmektedir. Coğrafi-mekansal veritabanları tipik olarak çok geniştir ve veriler homojen değildir. Bundan dolayı Coğrafi-mekansal verilerin sadece geleneksel raster ve vektör formatlarda değil aynı zamanda yapılandırılmamış ve yarı yapılandırılmış verileri içeren (multimedya verileri gibi) yeni biçimlerinin işlenebilmesi için veri madenciliği metotları geliştirilmiştir.

MVM'nin temel görevi bazı gizli, bilinmeyen ve ilginç mekansal/zamansal örüntü dağılımlarını ayırt etmek ve çıkarmaktır. Coğrafi-mekansal veri analizi için güncel metotlar mekansal kümeleme algoritmaları, sapma analizi metotları, mekansal sınıflama ve birliktelik analizi ile görsel veri madenciliği metotlarını kapsamaktadır.

MVM, mekansal verilerin analizi için temel yeni bir anlayış sağlar. Genellikle veri madenciliğinin heterojen coğrafi verilerin birleştirilmesini gerektiren analizi için pek çok uygulamalar vardır. Örneğin, çiftçilerin verimlilik değerlendirmesi, su kalitesi analizi, mal depolama dağılımı, suç veya hastalık kontrolü (Campbell, James, 1996, Jasani, Stein, 2002, Swain, Davis, 1978). Çevresel sağlık tehlikeleri araştırmaları için kanser noktalarının kümelemeleri, polis devriye rotaları planlaması için suç sıcak noktalarının tespiti, kuş gribi virüsünün yayılma noktaları, küresel ısınmanın hava durumunu etkilediği bölgelerin tespiti verilebilecek örneklerden sadece birkaçıdır (Shekhar, Chawla, 2003). Yine bir arazi gözlem veritabanından benzer toprak kullanım alanlarının tanımlanması, şehir planlamada ev tiplerine, değerlerine ve coğrafi konumlarına göre ev gruplarının tanımlanması da MVM ile belirlenebilen özelliklerdir. Buradan çıkarılacak bilgiler ekolojik ve çevresel yönetim, halk güvenliği, ulaşım, halk sağlığı, ticari, tarım, endüstri, seyahat, turizm, ulusal savunma vs. konularda uzman kararlarının alınmasını kolaylaştıracaktır.



Şekil 3: Birleştirilmiş sistem: sağ pencere interaktif görsel veri analizi araçları ile bir haritayı göstermektedir. Sol pencere tematik verilere karar ağaçları uygulamasının sonuçlarını göstermektedir (Andrienko ve diğ., 2001).

Mekansal veritabanlarından (örneğin CBS) bilgi keşfi; ilginç, faydalı ve beklenmedik mekansal örüntülerin ve özelliklerin çıkarımı, mekansal ve mekansal olmayan veriler arasında var olan genel ilişkiler ve mekansal veritabanlarında var olmayan diğer veri karakteristikleri ile ilişkilidir. MVM, kullanılmakta olan veritabanları gibi mekansal verinin anlaşılmasını ve gözden geçirilmesini, mekansal veri parçaları ve mekansal-mekansal olmayan parçalar arasındaki ilişkilerin açılmasını kapsar. Mekansal veri tabanlarında var olabilen gizli özellikler ve ilginç ilişkileri keşfeder. Bu uygulamalar;

- astronomi,
- çevresel analiz,
- uzaktan algılama,
- tıbbi görüntüleme,
- navigasyon vs.

gibi alanlarda kullanılır (Yubin, 2005).

Mekansal veri manuel analiz için çok hızlı şekilde büyümektedir. Uydu görüntüleri, GPS gibi gözlem verilerinin çok yoğun elde edildiği günümüzde bu verilerden en kısa zamanda bilgiye ulaşmakta çok önemli hale gelmiştir. Örneğin, uydu görüntüleri verileri ile sınıflama ağaçlarını kullanarak bitki örtüsü değişimleri çok hızlı bir şekilde gözlenebilmektedir. Bu işlemlerde MVM, kullanılabilir hipotez sayısını azaltabilir. Veriler tarafından desteklenen hipotezlerin tanımlanmasını sağlar.

Coğrafi Bilgi Teknolojileri (CBS) ve veri tabanlarından bilgi keşfi aralarında zayıf bağlantılar olan iki farklı teknoloji olarak geliştirilmiştir. Son zamanlarda kurumlar çok büyük miktarda coğrafi bilgiyi veritabanlarında toplamış ve burada gizli bulunan bilginin potansiyelini fark etmeye başlamışlardır. Bu bakımdan CBS'ye veri madenciliği teknolojilerinin uygulanması işlemi oldukça zamanında ve önemli olmuştur. Günümüzde pek çok CBS ve veri madenciliği sistemlerinin birleştirilmesi ile inşa edilmiş tümleşik sistemler mevcuttur. Bu şekilde veri madenciliği araçlarıyla otomatikleştirilmiş kartografik görselleme ve grafik gösterimlerin birbirleriyle etkileşimli olarak işlenmesi sağlanacaktır ve mekânsal verinin analiz işleminin verimi daha da artacaktır (Andrienko ve diğ., 2001). Pek çok MVM araçları CBS araçları ile etkileşimli çalışabilmektedir. Böyle bir sistemin ara yüzü şekil 3'te gösterilmiştir. Mekânsal istatistik metotları mekansal veri madenciliğinin bir parçasıdır. Bu metotlardan bazıları CBS ve analiz araçlarında uygulanmaktadır.

2. SONUÇLAR

Veri toplama araçları ve veri tabanı teknolojilerindeki gelişme, veritabanlarında ve diğer bilgi depolarında çok miktarda bilgi depolanması sonucunu doğurmuş tam anlamıyla bir veri patlaması yaşanmıştır. Veri madenciliği bu veritabanlarındaki büyük miktarda ham veriden ilginç örüntülerin ve değerli bilginin keşfini sağlayan bir metotlar bütünüdür. Kullanılacak metot istenilen bilgiye ve eldeki veriye göre değişmektedir. Çözümün kalitesi öncelikle verinin kalitesine bağlıdır. Veri madenciliği, uygulama alanındaki uzmanların ve bilgisayarın ortak çalışmasıdır.

2000'li yıllarda dünya gündeminde veri madenciliği oldukça popüler bir hale gelmiştir. Veritabanlarındaki tüm verinin % 80'inin mekânsal bileşenlere sahip olduğu düşünülürse mekânsal verilerden bilgi keşfi içinde mekansal veritabanlarından veri madenciliği uygulamalarının ne derece faydalı olacağı anlaşılabilir.

Son zamanlarda kurumlar çok büyük miktarda coğrafi bilgiyi veritabanlarında toplamış ve burada gizli bulunan bilginin potansiyelini fark etmeye başlamışlardır. Bu bakımdan CBS'ye veri madenciliği teknolojilerinin uygulanması işlemi oldukça zamanında ve önemli olmuştur.

KAYNAKLAR

ANDRIENKO, G. L.; ANDRIENKO, N. V.; SAVINOV, A. A. (2001), *Integration of GIS with Data Mining*, Statistical Office of the European Communities, ISBN: 92-828-7883-X S.221-225.

AYDOĞAN, F. (2003), *E-Ticarette Veri Madenciliği Yaklaşımlarıyla Müşteriye Hizmet Sunan Akıllı Modüllerin Tasarımı ve Gerçekleştirmi*, Yüksek Lisans Tezi, Hacettepe Üniversitesi.

BAYKAL, N. (2003), Tıp Bilişimi Güz Okulu Semineri.

BAYKASOĞLU, A. (2005), "Veri madenciliği ve çimento sektörüne bir uygulama", *Akademik Bilişim Konferansı, Gaziantep Üniversitesi, Gaziantep*, (82-83).

CAMPBELL, JAMES B. (1996), *Introduction to Remote Sensing*, The Guilford Press, New York.

CASTRO, V. E.; LEE, I. (2004), "Clustering with obstacles for Geographical Data Mining", *ISPRS Journal of Photogrammetry & Remote Sensing,* vol. 59, (21– 34).

EKER, H. (2007), "Veri Madenciliği Veya Bilgi Keşfi", http://www.bilgiyonetimi.org/cm/ (10.09.2007).

ESTER, M.; KRİEGEL, H. P.; XU, X. (1995), "Knowledge Discovery in Large Spatial Databases: Focusing Techniques for Efficient Class Identification", *Proc. of the Fourth Int. Symposium on Large Spatial Databases (SSD '95)*, Portland, Maine, USA

JASANİ, B.; STEİN, G. (2002), Commercial *Satellite Imagery, a tactic in nuclear weapon deterrence*, Springer.

MİLLER, H. J.; HAN, J. (1999), "Discoverıng geographic knowledge in data rich environments", *Report of a Specialist Meeting held under the auspices of the Varenius Project*, Kirkland, Washington

OPENSHAW, S. (1999), "Geographical data mining: key design issues", *Proceedings of the 4th GeoComputation 99.*

SHEKHAR, S.; CHAWLA, S. (2003), *Spatial Databases: A Tour*, Prentice Hall, ISBN 013-017480-7.

SWAİN, P.H.; DAVİS, S.M. (1978), *Remote Sensing: the Quantitative Approach*, McGraw-Hill, Inc.

VAHAPLAR, A.; İNCEOĞLU, M. M. (2001), "Veri Madenciliği ve Elektronik Ticaret", *VII. Türkiye'de İnternet Konferansı.*

YUBIN, Y. (2005), Spatial Data Mining**,** *Lecture Notes*, Joint Laboratory for Geoinformation Science. The Chinese University of Hong Kong.

WANG, L.; XİE, K.; CHEN, T.; MA, X. (2005), "Efficient discovery of multilevel spatial association rules using partitions", *Information and Software Technology,* vol. 47, (829–840).

WU, H. C.; LU, C. N. (2002), "A Data Mining Approach for Spatial Modeling in Small Area Load Forecast", *IEEE Transactıons on Power Systems*, vol. 17, no. 2.