



VERİ MADENCİLİĞİNİN GÜVENLİK UYGULAMALARI VE KİŞİSEL GİZLİLİK

Hayrettin Bahşı

bahsi@uekae.tubitak.gov.tr

6 HAZİRAN 2008

- Veri Madenciliği Nedir?
- Veri Madenciliği Yöntemleri
- Veri Madenciliği Uygulamaları
- Veri Madenciliğinin Güvenlik Alanındaki Uygulamaları
- Kişisel Gizlilik ve Veri Madenciliği

- Bilgi sistemlerinin yaygınlaşması
- Bilgi sistemlerinde toplanan büyük miktarda veriler
 - Satır sayısı
 - Kolon sayısı
- Veri Madenciliği
 - Bilgi içerisinde yararlı kalıplar bulma
 - Geleceğin, gidişatın tahmini
 - Normal sorgularla bulunamayacak çıkarımlar

- Veri Madenciliği
 - İstatistik
 - Makine öğrenmesi
 - Yapay zeka
 - Veritabanı yönetimi
 - Görselleştirme
- Bilgi kaynağı
 - Yapısal
 - Veri tabanları
 - Yapısal olmayan
 - Yazılar, video, ses veya resim bilgileri, e-postalar

- Kullanılan yöntemlere örnekler
 - Yapay sinir ağları (Neural Networks)
 - Karar ağaçları (Decision Trees)
 - Bayesian ağlar (Bayesian Networks)
 - Kendi kendine organize olan ağlar (Self Organizing Maps)
- Son zamanlardaki veri madenciliği uygulamalar
 - Yapısal olmayan bilgiler üzerindeki uygulamalar
 - Uzmanlara yardımcı olacak görsel bileşenler
 - Semantik ağlar, sosyal ağlar

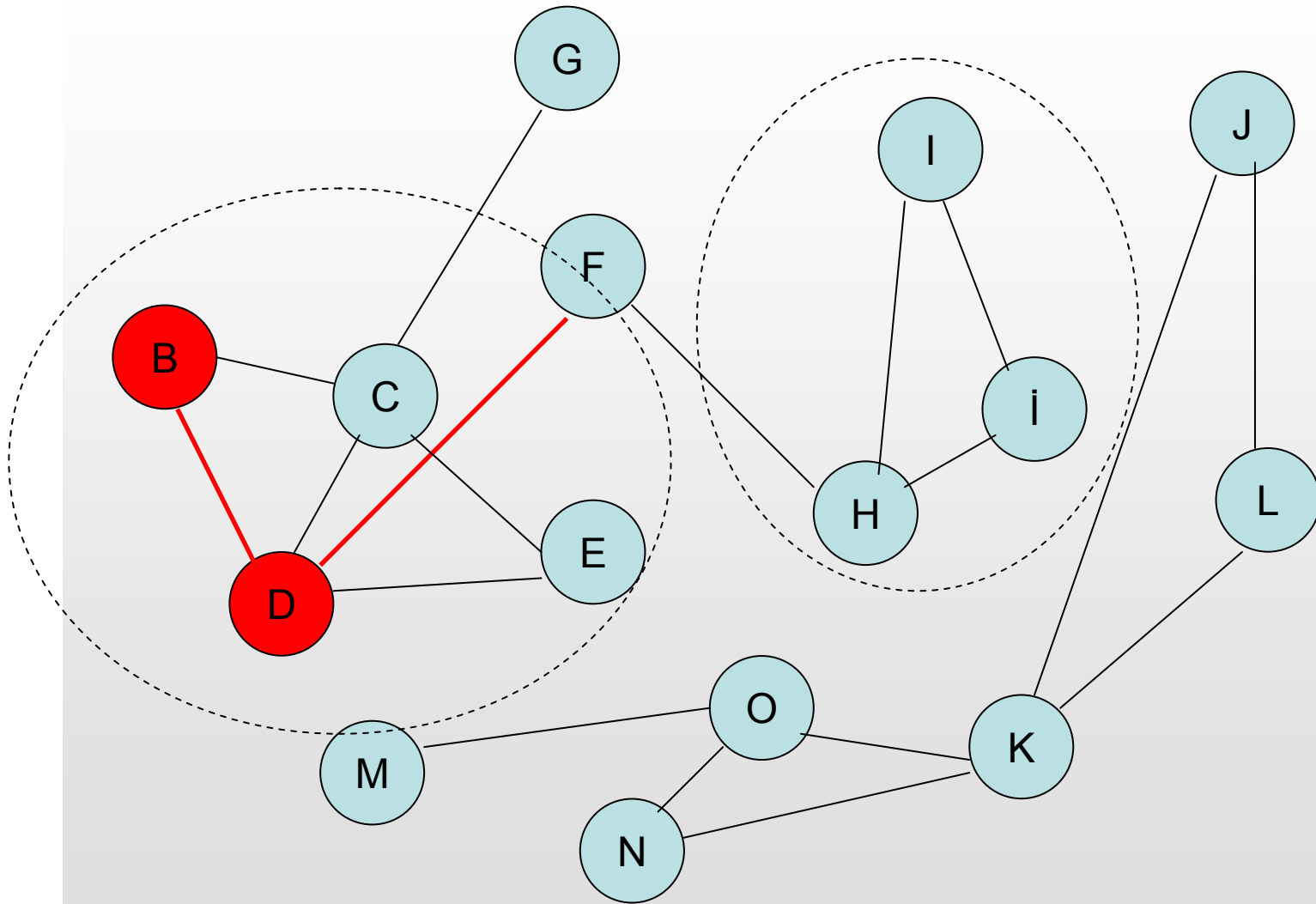
- Özel sektörde
 - Yeni iş imkanlarının belirlenmesi
 - Müşteri ihtiyaçlarının belirlenmesi
 - Satış trendlerinin belirlenmesi
 - İş performansının optimize edilmesi
- Kamu sektöründe
 - Vergi politikalarının belirlenmesi
 - Sahtekarlık ve suistimallerin tespit edilmesi
 - Kara para aklama, vergi kaçakçılığı

- Dolandırıcılık Tespiti
 - Şüpheli aktivitelerin belirlenmesi
 - Analiz yapacak kişilere şüpheli aktivitelerin aktarılması
 - Bazı alanlar
 - Sigorta dolandırıcılıkları
 - Kredi dolandırıcılıkları
 - Vergi dolandırıcılıkları
 - Kara para aklama faaliyetleri
 - Kurum içi dolandırıcılıklar
- Terörist/suç gruplarının tespiti
- Bilgi sistemleri güvenliği
 - Spam e-postalarının tespiti
 - Anomali tabanlı saldırı tespit sistemleri

- Bir banka
 - Tüm müşteri bilgilerinin analiz edilmesi
 - Kredi taksitlerini ödemeyen müşterilerin profillerinin belirlenmesi
 - Kredi talep eden kişinin bilginin eldeki profillere uygunluğu
 - Kredi verip vermeme kararı

- İletişime ait bilgiler
 - Telefon konuşmaları
 - E-postalar
- Suç kayıtları
- Tüm bilgilerin ağ topolojisine dökülmesi
- Her kişinin ağda bir nod olması
- Linklerin kişiler arasındaki ilişkileri belirlemesi
- Linklerin özellikleri, ağırlıkları
- Ağ içerisindeki yoğun kümelerin bulunması

Terörist/Suç Gruplarının Tespiti (2)



- A.B.D'nin Yurt Güvenliği (Homeland Security) Kurumunun veri madenciliğine verdiği önem
 - “Survey of DHS Data Mining Activities”
 - Veri madenciliğinin kullanıldığı sistemler
 - United States Customs and Border Protection
 - Immigration and Customs Enforcement
 - Office of Intelligence and Analysis
 - United States Secret Service
 - Transportation Security Administration

- ABD'deki kurum ve kuruluşların net karının %5 i mali sahtekarlık sebebiyle kaybediliyor. Kayıp yılda 652 milyar \$ (ACFE)
- Avustralya'da kredi kartı sahtekarlığında dönen işlem yıllık mali işlemin %0.0148 ini oluşturuyor.
- ABD de sadece 1998 yılında 17 federal programda toplam 19.1 milyar \$ kaybın mali sahtekarlık yüzünden olduğu
- 19.1 milyar \$ toplamın 12 milyarlık kısmının sağlık sektöründe olduğu biliniyor.(GAO)

- Credit Card Issuer Bank 2003 yılında kurduğu veri madenciliği çözümleriyle ilk on ayda 3 milyon \$ sahtekarlık saptaması yapmış ve önlemiştir.
- HSBC Bank, veri madenciliği yöntemleriyle
 1. Potansiyel sahtekarlık işlemlerinin tespitinde % 10 artış
 2. Tespit sırasında harcanan işlemci gücünde % 12 azalma
 3. Tespit için incelenen işlem-hesap miktarında % 87 artış elde etmiştir.

- Medical Benefits Fund (Avustralya) 2000 yılında kurduđu veri madenciliđi çözümlü ile veri madenciliđi alt yapısına harcadıđı paranın 2 katını ilk 6 ayda sahtekarlıktan kurtardıđını duyurdu.
- Leòn Bank (Dominik Cumh.) mali sahtekarlık kaynaklı para kayıplarını üç ayda %30 azalttıđını duyurdu.

- Kişisel gizlilik: Bir organizasyonun ya da kişinin kendisine ait bilginin akışını kontrol edebilmesi
- Kişiyi ya da kurumu ilgilendirebilir
- “Kimin neyi bilmesi gerektiği” sınırlarının kanunlarla belirlenmesi gerekliliği
- Veri madenciliği uygulamalarının çok miktarda veriye erişebilmesi
- Verinin işlenmesi sonucu çıkan çıkarımların da kişiye özel olabileceği

- Terörle mücadelede veri analizi-madenciliği kullanılması
 - 11/09 sonrası
 - Özellikle ABD’de
- ABD’de yürütülen ADVISE projesi
 - Detaylı kişisel bilgileri barındıran hacimli veri ambarlarını işlemek
 - Terörist ağları keşfetmeye yönelik
 - Kişisel verilerin güvenli ortamlarda saklanması-işlenmesi sorunu

- Kişisel gizliliğin yoğun tartışılması sonucu:
 - i. Pentagon TALON isimli anti-terörizm programını durdurmak zorunda kaldı.
 - ii. Kongreye bağlı TIA(Tümleşik Enformasyon Bilinci) projesi askıya alındı. (2003)
 - iii. Yurt Güvenliği Birimi'nin (DHS) 42 milyon\$ harcadığı ADVISE projesinde test aşamasında gerçek veri kullanıldığı GAO raporlarında saptanınca süresiz iptal edildi.

- Bireysel gizliliğin korunması için:
 - “Bireysel gizlilik” ile “işe yarar çıkarımlar” dengesi
 - Mümkün oldukça gerçek veri özelliklerine yakın gerçek olmayan verinin sistemlerin eğitimi sırasında kullanılması
 - Bilmesi gereken prensibine göre bilgiye erişim
 - Bilginin kişisel gizliliğe ait detaylardan arındırılması
 - Erişimlerin, sorguların, veri madenciliği aktivitelerinin denetim kayıtlarının oluşturulması

- Ülkemiz ve dünya bilgi toplama ve depolama evresinde
- Yakın gelecekte bilginin veri madenciliği ile işlenmesi yaygınlaşacak
- Kamu kurumları
 - Sahtecilik ve suistimal tespitlerinde
 - Stratejik kararlardaveri madenciliği yöntemlerini kullanabilir.

- Kişisel gizlilik, veri madenciliğinin en büyük probleminden birisi
- Kişisel gizlilik düşünülerek etkin veri madenciliği yöntemleri uygulanmalı
 - Proje iptallerinin (kaynak israfının) önlenmesi
 - Hukuki davalara maruz kalmama



Teşekkürler Sorular?

Hayrettin Bahşı

bahsi@uekae.tubitak.gov.tr