

Sıra Desen Madenciliği

Sıra desen madenciliği, belirli bir sırayla gelen istatistiksel olarak veri örnekleri arasındaki ilgili örüntüleri bulmaya çalışır. Müşterilerin teknoloji market alışverişi verilerine göre son 3 ayda sırasıyla önce bilgisayar sonra CD-ROM son olarak dijital kamera satın almaları, tıbbi tedaviler, doğal felaketler(deprem), DNA dizilişi ve gen yapısı sıralı örüntü madenciliği ile ilgilidir. Mesela İnternet şubesinde yapılan işlemler bir sıraya göre yapıldığı için sıralı örüntü madenciliği içerisinde yer almaktadır. sıralı örüntü madenciliğinde arka arkaya yapılan işlemler göz önüne alınır.

PrefixSpan:

PrefixSpan (Prefix-projed Sequential pattern mining) algoritması, sıralı veri madenciliği için çok iyi bilinen bir algoritmadır. Ardışık desenleri, desen büyütme yöntemi ile çıkarır. Algoritma, küçük veri kümeleri için çok iyi performans gösteriyor.

PrefixSpan sık tekrarlanan işleri tekrar eder. Örnek vererek anlatmak gerekirse bir sürecin belirli zaman dilimlerinde nasıl tekrarlandığını inceler. Bu tekrarın hangi işlemler öncesi veya sonrası olduğuda dikkate alınmaktadır Algoritmaya yönelik manipilasyon bulmada kritik öneme sahiptir. Süreç analizinde gelecekte oluşacak krizlerin izlerini aramada önem katmakta.

Örnek: Aşağıdaki tabloda belli eylemler ve bu eylemlerin zaman dilimlerine göre gerçekleşme akışları verilmektedir. Veri madenciliğinin önemli adımlarından biri olan veri etiketleme süreci bu algoritmanında akışında rol oynamaktadır.

Burada A – B – C eylemlerini bir internet bankacılığı süreçlerindeki havale(A), ödeme(B), giriş işlemleri(C) gibi etiketleme yaptıktan sonra süreci bir ID tanımlayarak sayısallaştırmaktadırlar. Bu sayıda algoritmanın temelinde yatan örüntü yakalama amacınıda hızlandırmış olmaktadır. Burada genel yaklaşımı anladıktan sonra, gerçek bir örnek üzerinde ilerleyebiliriz.

Object	Timestamp	Events
A	10	2, 3, 5
A	20	6, 1
A	23	1
B	11	4, 5, 6
B	17	2
B	21	7, 8, 1, 2
B	28	1, 6
C	14	1, 8, 7

Aşağıdaki şekil de görüldüğü gibi sıralı örüntü madenciliği yöntemine göre müşterilerin herhangi bir t anında sırayla yaptığı işlemleri, satın aldığı ürünleri, ziyaret ettiği web sayfalarını analiz ederek müşterinin davranışına uygun çıkarımlar yapılabilir. Bir sıralı veri analizi uygulamasının başarılı olması için destek (support) değerinin en küçük destek değerinden (minsupport) başka ifade ile eşik değerinden daha büyük olması gerekir.

