

APRIORI ALGORİTMASININ FARKLI VERİ KÜMELERİNE UYGULANMASI

M. Emin Eker

Veri Madenciliđi

Daha 6nceden bilinmeyen, ge6erli ve uygulanabilir bilgilerin geniř veritabanlarından elde edilmesi ve bu bilgilerin iřletme kararları verirken kullanılmasıdır.

řu genel ama6lar i6in kullanılır:

- Sınıflandırma
- Birliktelik Kuralları ve İliřki Analizi
- K6meleme

Birliktelik Kuralları

Birliktelik kuralı, geçmiş verilerin analiz edilerek bu veriler içindeki birliktelik davranışlarının tespiti ile geleceğe yönelik çalışmalar yapılmasını destekleyen bir yaklaşımdır.

Birliktelik kurallarının kullanıldığı en tipik örnek Market Sepeti Analizi'dir (Pazar Sepeti Analizi).

Market Sepet Analizi

Bir müşteri süt satın alıyorsa, aynı alışverişte sütün yanında ekmek alma olasılığı nedir? Bu tip bir bilgi ışığında rafları düzenleyen market yöneticileri ürünlerindeki satış oranını arttırabilirler. Örneğin bir marketin müşterilerinin süt ile birlikte ekmek satın alma oranı yüksekse, market yöneticileri süt ile ekmek raflarını yan yana koyarak ekmek satışlarını arttırabilirler.

Örneğin; bir A ürününü satın alan müşteriler aynı zamanda B ürününü de satın alıyorsa, bu durum Birliktelik Kuralı ile gösterilir.



Algoritmalar

AIS

SETM

Apriori

AprioriTid

Apriori-Hybrid

OCD (Off-line Candidate Determination)

Partitioning

CARMA (Continuous Association Rule Mining)

Count Distribution

Intelligent Data Distribution

Parallel Association Rules

...

Apriori Algoritması

Apriori algoritması, Agrawal ve Srikant tarafından 1994 yılında geliştirilmiştir. [IBM]

Veri Madenciliğinde, birliktelik kuralı çıkarım algoritmaları içerisinde en fazla bilinen ve kullanılan algoritmadır.

Algoritmanın ismi, yaygın nesnelerin önsel bilgilerini kullanmasından yani bilgileri bir önceki adımdan almasından “önceki (prior)” kelimesinden gelmektedir.

Temel Yaklaşım

Öğe kümesi (itemset)

Bir veya daha çok öğeden oluşan küme

k-öğelik kümesi (k-itemset): k öğeden oluşan küme

3-öğelik kümesi: {Bal, Süt, Ekmek}

Bu algoritmada temel yaklaşım, “eğer k-öğelik kümesi minimum destek kriterini sağlıyorsa, bu kümenin alt kümeleri de minimum destek kriterini sağlar.” şeklindedir.

Destek ve Güven Kriterleri

Birliktelik kuralında, öğeler arasındaki birliktelik, destek ve güven kriterleri ile hesaplanır. Destek (Support) kriteri, veride öğeler arasındaki bağıntının ne kadar sık olduğunu belirtir.

X ve Y farklı ürünler olmak üzere,

X ürünü için destek, tüm alışverişler içinde X ürününün oranıdır.

$$\text{DESTEK}(X) = X \text{ Sayısı} / \text{Toplam Alışveriş Sayısı}$$

X ve Y ürünleri için destek, X ve Y'nin bir arada tüm alışverişler içinde bulunma olasılığıdır.

$$\text{DESTEK}(X, Y) = (X, Y) \text{ Sayısı} / \text{Toplam Alışveriş Sayısı}$$

Destek ve Güven Kriterleri

Güven (confidence) kriteri ise Y ürününün hangi olasılıkla X ürünü ile beraber olacağını söyler.

$$\text{GÜVEN}(X, Y) = (X, Y) \text{ Sayısı} / X\text{'i İçeren Alışveriş Sayısı}$$

$$\text{GÜVEN}(X \Rightarrow Y) = \text{DESTEK}(X, Y) / \text{DESTEK}(X)$$

Elde edilen kuralların güvenilirliği, destek ve güven değerleri ile doğru orantılıdır.

Destek ve Güven Kriterleri

Her kural bir destek ve güven değeri ile ifade edilir.

$A \Rightarrow B$ [destek = 2%, güven = 60%]

Birliktelik kuralı için 2% destek değeri, analiz edilen tüm alışverişlerden 2%'sinde A ile B ürünlerinin birlikte satıldığını belirtir.

60% oranındaki güven değeri ise A ürününü satın alan müşterilerin 60%'inin aynı alışverişte B ürününü de satın aldığını gösterir.

Algoritmanın Adımları

1. Minimum destek sayısı (min.support) ve minimum güven değerinin (min.confidence) belirlenmesi
2. Öge kümeler içerisindeki her bir ögenin destek değerinin bulunması
3. Minimum destek değerinden daha düşük desteğe sahip olan ögelerin devre dışı bırakılması
4. Elde edilen tekli birliktelikler dikkate alınarak ikili birlikteliklerin oluşturulması
5. Minimum destek değerinden düşük olan öge kümelerinin çıkarılması
6. Üçlü birlikteliklerin oluşturulması
7. Üçlü birlikteliklerden minimum destek değerini geçenlerin dışındakilerin çıkarılması
8. Üçlü birlikteliklerden birliktelik kurallarının çıkarılması

Örnek

ID	SEPET
1	Elma, Muz, Dondurma, Simit
2	Elma, Muz, Simit
3	Yumurta, Simit
4	Yumurta, Erik
5	Elma, Muz
6	Elma, Muz, Yumurta

Yandaki verilere minimum destek %30 ve güven %60 olacak şekilde Apriori algoritmasını uygulayalım.

Örnek

Birinci Tarama

ÜRÜN	MİKTAR	DESTEK
ELMA	4	%67
MUZ	4	%67
DONDURMA	1	%17
SİMİT	3	%50
YUMURTA	3	%50
ERİK	1	%17

Örnek

İkinci Tarama

ÜRÜN	MİKTAR	DESTEK
ELMA, MUZ	4	%67
ELMA, SİMİT	2	%33
ELMA, YUMURTA	1	%17
MUZ, SİMİT	2	%33
MUZ, YUMURTA	1	%17
SİMİT, YUMURTA	1	%17

Örnek

Üçüncü Tarama

ÜRÜN	MİKTAR	DESTEK
ELMA, MUZ, SİMİT	2	%33

Örnek

- Elma alanlar (4 kayıt), Muz ve Simit alır (2 Kayıt) [Destek %33, Güven = %50]
- Muz alanlar (4 kayıt), Elma ve Simit alır (2 Kayıt) [Destek %33, Güven = %50]
- Simit alanlar (3 kayıt), Elma ve Muz alır (2 Kayıt) [Destek %33, Güven = %67]
- Elma ve Muz alanlar (4 kayıt), Simit alır (2 Kayıt) [Destek %33, Güven = %50]
- Elma ve Simit alanlar (2 kayıt), Muz alır (2 Kayıt) [Destek %33, Güven = %100]

Yukarıdaki sonuçlara göre minimum destek %30 ve güven %60 şartını sağlayan sadece iki adet birliktelik kuralı elde edilmiştir.

- Simit alanlar (3 kayıt), Elma ve Muz alır (2 Kayıt) [Destek %33, Güven = %67]
- Elma ve Simit alanlar (2 kayıt), Muz alır (2 Kayıt) [Destek %33, Güven = %100]

Örnek

Simit alanlar (3 kayıt), Elma ve Muz alır (2 Kayıt) [Destek %33, Güven = %67]

Elma ve Simit alanlar (2 kayıt), Muz alır (2 Kayıt) [Destek %33, Güven = %100]

ÜRÜN	MİKTAR	DESTEK
ELMA, MUZ, SİMİT	2	%33

Elimizdeki veritabanına göre Elma – Muz – Simit en geniş nesne kümesidir. Güven seviyesi ise bakış açısına göre değişmektedir. Örneğin her elma ve simit alan, mutlaka muz da satın almıştır; bununla beraber her elma ve muz alan, mutlaka simit almıştır diyemiyoruz.

Uygulama Alanları

Mühendislik

Tıp

Eğitim

Bankacılık

Finans

Telekomünikasyon

Pazarlama

E-Ticaret

Sigortacılık

...

Veri Madenciliđi Programları

Veri Madenciliđi uygulamalarını gerekleřtirmek iin programlara ihtiya duyulur. Bu kapsamda;

Ticari: SPSS Clementine, SPSS, SAS, Angoss, KXEN, MATLAB

Aık kaynak: RapidMiner (YALE), WEKA, R, C4.5, Orange, KNIME

olmak üzere birok program geliřtirilmiřtir.

Örnek Uygulamaları

Ege Bölgesi'ndeki bir Araştırma ve Uygulama Hastanesinin Acil Hasta Verilerinin Veri Madenciliği ile Analiz Edilmesi [Makale] [Sabri ERDEM, Güzin ÖZDAĞOĞLU]

“Bu çalışma kapsamında, belirli bir dönem boyunca Ege Bölgesi'ndeki bir araştırma ve uygulama hastanesinin acil servisine başvuruda bulunan 214 bin hasta verisi ele alınarak, Veri Madenciliğinde sıklıkla kullanılan birliktelik kuralı yöntemiyle, veri setindeki gizli ancak anlamlılık içeren ilişkiler ortaya çıkarılmaya çalışılmıştır. Çalışma sonuçları, bölgesel özellikler taşıyabileceği düşünülen acil servislere hastaların başvuru nedenleri ve hasta profilleri açısından bir fikir vermekte, ayrıca acil servis bölümlerinin yeniden yapılanma çalışmalarına da farklı bir açıdan yol göstererek katkıda bulunmaktadır.”

Örnek Uygulamaları

ERP Yazılımlarının İşletmelerdeki Etkilerinin Tanımlayıcı Veri Madenciliği Yöntemleriyle İncelenmesi [Bildiri] [Ahmet Selman BOZKIR, Betül YAVAŞOĞLU]

“Bu çalışmada tanımlayıcı veri madenciliği yöntemlerinden birliktelik kuralları analizi ile kurumların ERP kullanmaları halinde kazançlarının ne olacağı, geçiş öncesinde ne gibi altyapısal unsurlara sahip olmaları gerektiği ortaya çıkarılmaya çalışılmıştır. Sonuçlara bakıldığında güncel kapasite kullanım oranlarının ve bazı departmanlara sahip olma yüzdesinin ERP sistemlere sahip olma ile yüksek derecede pozitif korelasyon gösterdiği tespit edilmiştir. Öte yandan, birliktelik kuralları analizi ile ERP sistemlere sahip olan ve olmayan işletmelerde sıklıkla gözlemlenen kurallar çıkarılmıştır.”

Örnek Uygulamaları

**Otomobil Yetkili Servislerinde Birliktelik Kurallarının
Belirlenmesinde Apriori ve FP-Growth Algoritmalarının
Karşılaştırılması** [Bildiri] [Semra Erpolat]

“Bu çalışmada Türkiye’de otomotiv sektöründe faaliyet gösteren bir yetkili servisin müşterilerine ait alışveriş verileri, Apriori ve FP-Growth Algoritmaları kullanılarak analiz edilmiştir. Böylelikle müşterilerin hangi ürünleri birlikte satın aldıkları gözlemlenmiş ve bu doğrultuda karı artırmaya yönelik uygulanacak kampanya va promosyonlara yön verilmeye çalışılmıştır.”

Örnek Uygulamaları

Birliktelik Kuralı Yöntemi için bir Veri Madenciliği Yazılımı Tasarımı ve Uygulaması [Bildiri] [Feridun Cemal ÖZÇAKIR, A. Yılmaz ÇAMURCU]

“Bu çalışmada, bir firmanın pastane satış verileri üzerinde veri madenciliği uygulamak için birliktelik kuralları ile bir yazılım tasarlanmıştır. Tasarlanan yazılımda, Apriori algoritması kullanılmıştır. Uygulanan Apriori algoritması ile farklı zaman dilimi, farklı satış lokasyonu girdi değerleri doğrultusunda birlikte satın alınan ürünler ile ilgili bağıntılar olduğu gözlemlenmiştir. Genelde aynı ürün grubuna ait ürünlerin, en sık birlikte satın alınan ürünler olduğu görülmüştür.”

Örnek Uygulamaları

MİGROS Türk A.Ş.'de Birliktelik Kurallarının Yerleşim Düzeni Planlamada Kullanılması [Makale] [Derya AY, İbrahim ÇİL]

“Bu çalışmada süpermarket yerleşim yeri düzenlemede veri yönelimli bir karar destek sistemi uygulaması sunulmaktadır. Çalışmada veri tabanlarında bilgi keşfi süreci kullanılarak yerleşim düzeni geliştirmeyi gerçekleştiren metodolojik bir çerçeve sunulmaktadır. Çalışmada ilişkisel veri tabanı hazırlanarak, Apriori algoritması ve Çok Boyutlu Ölçekleme (ÇBÖ) yöntemleri kullanılmaktadır. Deneysel çalışma ise Türkiye'nin önde gelen perakende satış firmalarından biri olan Migros Türk A.Ş.'de gerçekleştirilmektedir.”

Örnek Uygulamaları

Potansiyel Abonelerin Kazanılmasına Yönelik Pazarlama Stratejilerinin Veri Madenciliği İle Geliştirilmesi: Konya Şehri Doğalgaz Örneği [Yüksek Lisans Tezi] [Ali Yawar JAFARİ]

Ürün Kategorileri Arasındaki Satış İlişkisinin Birliktelik Kuralları ve Kümeleme Analizi İle Belirlenmesi ve Perekende Sektöründe Bir Uygulama [Doktora Tezi] [Ertuğrul ERGÜN]

Örnek Uygulamaları

Veri Madenciliğinde Apriori Algoritması ve Apriori Algoritmasının Farklı Veri Kümelerinde Uygulanması [Yüksek Lisans Tezi] [Ali Cenk GÜLCE]

“Bu tez çalışmasında, veri madenciliği ile ilgili kavramlar ve özellikle market sepet analizinde kullanılmak üzere birliktelik kuralları üreten apriori algoritması detaylı bir şekilde ele alınmıştır. Apriori algoritması market sepet analizinden farklı bir veri küme seti üzerine uygulanmıştır. Örnek anket veri setinden, apriori algoritması kullanılarak birliktelik kurallarını bulan bir uygulama geliştirilmiştir.”

İlgili Örnek Çalışmalar

- Mühendislik Öğrencilerinin Matematik I Derslerindeki Başarısının Destek Vektör Makineleri Kullanılarak Tahmin Edilmesi [Bildiri] [Necdet GÜNER, Emre ÇOMAK]
- Öğrenci Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi [Bildiri] [Çağdaş KURT, O. Ayhan ERDEM]
- Öğrencilerin Mezuniyet Notlarının Veri Madenciliği Metotları İle Tahmini [Bildiri] [Dönüş ŞENGÜR, Ahmet TEKİN]
- Apriori Algoritması ile Öğrenci Başarı Analizi [Bildiri] [Murat KARABATAK, Melih Cevdet İNCE]

İlgili Örnek Çalışmalar

Veri Madenciliği Algoritmalarını Kullanarak Öğrenci Verilerinden Birliktelik Kurallarının Çıkarılması [Yüksek Lisans Tezi] [Ufuk EKİM]

“Bu tez çalışmasında, halen Selçuk Üniversitesi’nde kullanılan öğrenci işleri otomasyonundan elde edilen veriler üzerinden, öğrenciler hakkında gelecekle ilgili tahmin yapılabilmesi için gerekli birliktelik kuralları çıkarılmıştır. Bu amaçla, bu tezde apriori algoritması ve karar ağacı algoritması kullanılmıştır. Bu kurallar sayesinde, Selçuk Üniversitesini yeni kazanan bir öğrencinin, üniversitedeki başarısına etki eden faktörler araştırılmıştır. Çalışma sonucunda, ailenin eğitim seviyesinin ve gelir düzeyinin öğrencinin başarısında en etkili faktörler olduğu görülmüştür.”

Uygulama

Farklı sınav kitapçıkları ile birlikte sınav oluşturan ve istatistiki olarak sınav analizi yapabilen bir eğitim yazılımı mevcuttur.

Sınavlar sonucunda oluşan sınav verileri 20 farklı kritere sahiptir. Bu verilere veri madenciliği adımları uygulanmıştır.

Apriori Algoritması ile yazılan bir eklenti hazırlanarak sisteme entegre edilmiştir. Bu sayede sistem üzerinde yapılan sınavların ardından veri madenciliği işlemi de uygulanarak çıkarımlar yapılmıştır. Bu çıkarımların ise hem sınav yöneticilerine hem de sınava katılanlara sunulması amaçlanmıştır.

Teşekkürler

