

VERİ MADENCİLİĞİ

(Birliktelik Kuralları)

Yrd.Doç.Dr. Kadriye ERGÜN
kergun@balikesir.edu.tr

İçerik

- Birliktelik Kurallarının Tanımı
- Destek ve Güven Ölçütleri
- Apriori Algoritması

Birliktelik Kuralları (Association Rules)

- Birliktelik kuralları
 - Veri kümesi içindeki yaygın örüntülerin, nesnelere oluşturan nitelikler arasındaki ilişkilerin bulunması □
- Birliktelik kurallarını kullanma: veri içindeki kuralları belirleme □
 - Hangi ürünler çoğunlukla birlikte satılıyor? □
 - Kişisel bilgisayar satın alan bir kişinin bir sonraki satın alacağı ürün ne olabilir? □
 - Yeni bir ilaca duyarlı olan DNA tipleri hangileridir? □
 - Web dokümanları otomatik olarak sınıflandırılabilir mi?

Destek ve Güven Ölçütleri

- Birliktelik çözümlerinin en yaygın uygulaması perakende satışlarda müşterilerin satın alma eğilimlerini belirlemek amacıyla yapılmaktadır. Müşterilerin bir anda satın aldığı tüm ürünleri ele alarak satın alma eğilimini ortaya koyan uygulamalara «pazar sepet çözümü» denir.
- Pazar sepet çözümlerinde satılan ürünler arasındaki ilişkileri ortaya koymak için «destek» ve «güven» gibi iki ölçütten yararlanır. Bu ölçütlerin hesaplanmasında destek sayısı adı verilen bir değer kullanılır. Kural destek ölçütü tüm alışverişler içinde hangi oranda tekrarlandığını belirler.

Destek ve Güven Ölçütleri

- Kural güven ölçütü A ürün grubunu alan müşterilerin B ürün grubunu da alma olasılığını ortaya koyar. A ürün grubunu alanların B ürün grubunu da alma durumu yani birliktelik kuralı $A \rightarrow B$ biçiminde gösterilir. Bu durumda kural destek ölçütü şu şekilde ifade edilir.

$$\text{destek}(A \rightarrow B) = \frac{\text{sayı}(A, B)}{N}$$

- Burada $\text{sayı}(A, B)$ destek sayısı A ve B ürün guruplarını birlikte içeren alışveriş sayısını göstermektedir. N ise tüm alışverişlerin sayısını göstermektedir. A ve B ürün gruplarının birlikte satın alınması olasılığını ifade eden kural güven ölçütü şu şekilde hesaplanır.

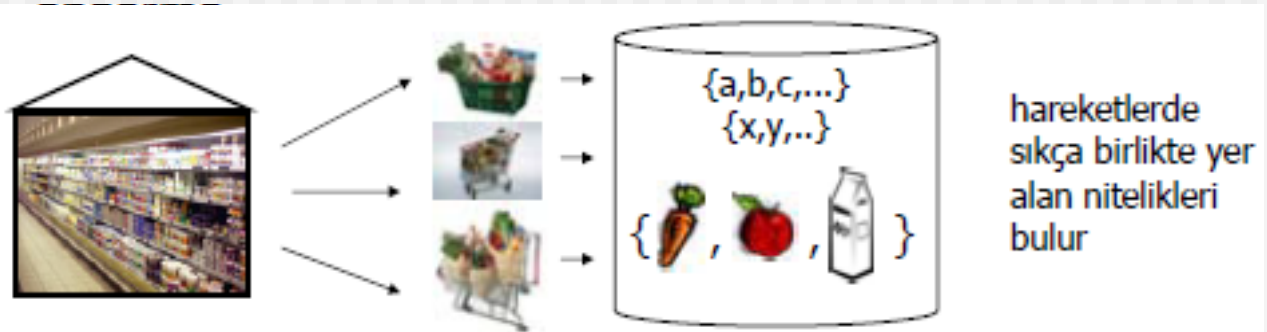
$$\text{güven}(A \rightarrow B) = \frac{\text{sayı}(A, B)}{\text{sayı}(A)}$$

Destek ve Güven Ölçütleri

- Birliktelik kuralları belirlenirken destek ve güven ölçütleri yanı sıra bu değerleri karşılaştırmak üzere eşik değere gereksinim vardır. Hesaplanan destek veya güven ölçütlerinin destek(eşik) ve güven(eşik) değerlerinden büyük olması beklenir. Hesaplanan destek veya güven ölçütleri ne kadar büyük ise birliktelik kurallarının da o derece güçlü olduğuna karar verilir.

Birliktelik Kuralları Bulma

- Bir niteliğin (veya nitelikler kümesinin) varlığını harekette bulunan başka niteliklerin varlıklarına dayanarak öngörme



- Kural şekli: **"Gövde → Baş [destek, güven] "**

satın alma(x, "ekmek") → satın alma (x, "süt") [%0.6, %65]

öğrenci(x, "BLG"), kayıt(x, "VTYS") → not(x, "A") [%1, %75]

Birliktelik Kuralları Bulma

- Bütün niteliklerden oluşan küme $I = \{i_1, i_2, \dots, i_d\}$
 - $I = \{\text{ekmek, süt, bira, kola, yumurta, bez}\}$
- Hareket $T_j \subseteq I, T_j = \{i_{j1}, i_{j2}, \dots, i_{jk}\}$
 - $T_1 = \{\text{ekmek, süt}\}$
- Hareketlerden oluşan veri kümesi $D = \{T_1, T_2, \dots, T_N\}$



Market Alışveriş verisi

Hareket	Öğeler
T1	Ekmek, Süt
T2	Ekmek, Bez, Bira, Yumurta
T3	Süt, Bez, Bira, Kola
T4	Ekmek, Süt, Bez, Bira
T5	Ekmek, Süt, Bez, Kola

Yaygın nitelikler:

Bez, bira
Süt, ekmek, yumurta, kola
Bira, ekmek, süt

Bulunan İlişkilendirme Kuralları

$\{\text{Bez}\} \rightarrow \{\text{Bira}\},$
 $\{\text{Süt, Ekmek}\} \rightarrow \{\text{Yumurta, Kola}\},$
 $\{\text{Bira, Ekmek}\} \rightarrow \{\text{Süt}\}$

Apriori Algoritması

- Birliktelik kurallarının üretilmesi için birçok yöntem kullanılmaktadır. Bunlardan en yaygın kullanılanı Apriori Algoritmasıdır.
- Apriori algoritması, özellikle çok büyük ölçekli veri tabanları üzerindeki veri madenciliği çalışmalarında geliştirilmiştir. Genel anlamda ilişki kuralı (association rule, birliktelik kuralı) çıkarımında kullanılan bir algoritmadır. Algoritmanın amacı, veri tabanında bulunan satırlar arasındaki bağlantıyı ortaya çıkarmaktır.
- Algoritmanın ismi, kendinden önceki çıkarımlara bağlı olduğu için, latince, önce anlamına gelen "prior" kelimesinden gelmektedir.
- Algoritma yapı olarak, aşağıdan yukarıya (bottom-up) yaklaşımı kullanmakta olup her seferinde tek bir elemanı incelemekte ve bu elemanla diğer adayların ilişkisini ortaya çıkarmaya çalışmaktadır.
- Ayrıca algoritmanın her eleman için çalışmasını, bir arama algoritmasına benzetmek mümkündür. Algoritma, bu anlamda genişlik öncelikli arama (breadth first search) yapısında olup adayları birer ağaç (tree) gibi düşünerek bu ağaç üzerinde arıyor kabul edilebilir.

Apriori Algoritmasının Adımları

- 1. Minimum destek sayısı ve minimum güven değerinin belirlenmesi
- 2. Öğe kümeler içerisindeki her bir ögenin destek değerinin bulunması
- 3. Minimum destek değerinden daha düşük desteğe sahip olan öğelerin devre dışı bırakılması
- 4. Elde edilen tekli birliktelikler dikkate alınarak ikili birlikteliklerin oluşturulması
- 5. Minimum destek değerinden düşük olan öğe kümelerinin çıkartılması
- 6. Üçlü birlikteliklerin oluşturulması
- 7. Üçlü birlikteliklerden minimum destek değerini geçenlerin dışındakilerin çıkarılması
- 8. Üçlü birlikteliklerden birliktelik kurallarının çıkarılması

Örnek 1.

- Bir mağazada alışveriş yapan müşterilere ilişkin olarak kayıtlar tutulmuş ve beş müşterinin yaptığı alışveriş göz önüne alınmıştır. Müşterilerin bir defada yaptığı alışverişler bir satırda yer almaktadır ve aşağıdaki tabloda verilmiştir. Bu tablodaki veriler kullanılarak müşteri davranışları Apriori Algoritmasıyla ortaya konmak isteniyor.

Müşteri	Aldığı Ürünler
1	Şeker, Çay, Ekmek
2	Ekmek, Peynir, Zeytin, Makarna
3	Şeker, Peynir, Deterjan, Ekmek, Makarna
4	Ekmek, Peynir, Çay, Makarna
5	Peynir, Makarna, Şeker, Bira

Örnek 1.

- a) Çözümlemeye bazı varsayımlarla başlanır. Destek ve güven ölçütleri için eşik değerleri belirlenir.

$$\text{destek(eşik)} = \%60$$

$$\text{güven(eşik)} = \%75$$

- Burada destek(eşik)=%60 olduğuna ve tüm müşteri sayısı 5 olduğuna göre **eşik destek sayısının** $(0,60)*5=3$ olduğu anlaşılır.
- b) Beş müşterinin alışveriş yaptığı ürünlerin kümesi {şeker, çay, ekmek, makarna, peynir, deterjan, bira, zeytin} biçimindedir. Nu ürünlerin her biri için destek değerleri hesaplanır.

$$\text{sayı(Şeker)}=3$$

$$\text{sayı(Çay)}=2$$

$$\text{sayı(Ekmek)}=4$$

$$\text{sayı(Makarna)}=4$$

$$\text{sayı(Deterjan)}=1$$

$$\text{sayı(Bira)}=1$$

$$\text{sayı(Zeytin)}=1$$

Örnek 1.

- Destek değerlerinin hesaplanması

Ürün	Sayı
Şeker	3
Çay	2
Ekmek	4
Makarna	4
Peynir	4
Deterjan	1
Bira	1
Zeytin	1

Örnek 1.

- c) Bu tablo üzerinde bazı ürünler eşik değere göre çıkarılır. Eşik destek sayısı 3 olduğuna göre bu eşik değerden küçük desteğe sahip olan ürünler çözümlenmeden çıkarılır. Buna göre oluşan yeni tablo aşağıdadır.

Ürün	Sayı
Şeker	3
Ekmek	4
Makarna	4
Peynir	4

Örnek 1.

- d) Çözümlemeye katılacak ürünler bu şekilde belirlendikten sonra ikili gruplar oluşturarak bu grupların destek sayıları hesaplanır.

$$\text{sayı}(\text{şeker,ekmek})=2$$

$$\text{sayı}(\text{şeker,makarna})=2$$

$$\text{sayı}(\text{şeker,ekmek})=2$$

$$\text{sayı}(\text{şeker,peynir})=2$$

$$\text{sayı}(\text{ekmek,makarna})=3$$

$$\text{sayı}(\text{ekmek,peynir})=3$$

$$\text{sayı}(\text{makarna,peynir})=4$$

Örnek 1.

- İkili ürün gruplarının destek değerleri

Ürün	Sayı
Şeker,Ekmek	2
Şeker,Makarna	2
Şeker,Peynir	2
Ekmek,Makarna	3
Ekmek,Peynir	3
Makarna,Peynir	4

Örnek 1.

- e) Bu tablodan bazı ürünler eşik değerine göre çıkarılır. Buna göre,

Ürün	Sayı
Ekmek, Makarna	3
Ekmek, Peynir	3
Makarna, Peynir	4

Örnek 1.

- f) Çözümlemeye katılacak ürünler bu şekilde belirlendiğine göre bu ürünlerin üçlü gruplar oluşturulur.

sayı(ekmek,makarna,şeker)=1

sayı(ekmek,makarna,çay)=1

sayı(ekmek,makarna,peynir)=3

...

sayı(ekmek,peynir,şeker)=1

sayı(ekmek,peynir,deterjan)=1

...

sayı(makarna,peynir,şeker)=2

sayı(makarna,peynir,çay)=1

...

Örnek 1.

- Eşik destek sayısına göre kalan üçlü ürün grupları aşağıdadır.

Ürün	Sayı
Ekmek, Makarna, Peynir	3

- Bu aşamadan sonra birliktelik kuralları elde edilebilir.

Örnek 1.

$$\text{sayı}(A,B)=\text{sayı}(\text{ekmek},\text{makarna},\text{peynir})=3$$

$$\begin{aligned}\text{destek}(A \rightarrow B) &= \frac{\text{sayı}(\text{ekmek}, \text{makarna}, \text{peynir})}{N} \\ &= \frac{3}{5} = 0.6\end{aligned}$$

biçiminde kural destek ölçütü elde edilir. Bu destek ölçütü koşul olarak verdiğimiz eşik değerden küçük değildir. O halde bu nesne kümesini kullanabileceğimiz anlaşılır. Kural destek sayılarına bağlı olarak birliktelik kuralları türeterek bu kurallar için güven ölçütlerini elde edeceğiz.

Sonuç 1:

$$\begin{aligned}\text{güven}(\text{Ekmek}, \text{makarna} \rightarrow \text{peynir}) &= \frac{\text{sayı}(\text{Ekmek}, \text{makarna}, \text{peynir})}{\text{sayı}(\text{Ekmek}, \text{makarna})} \\ &= \frac{3}{3} = \%100\end{aligned}$$

Örnek 1.

Sonuç 2: $güven(Ekmek \rightarrow peynir, makarna) = \frac{sayı(Ekmek, makarna, peynir)}{sayı(Ekmek)}$

$$= \frac{3}{4} = \%75$$

Sonuç 3: $güven(peynir \rightarrow ekmek, makarna) = \frac{sayı(Ekmek, makarna, peynir)}{sayı(peynir)}$

$$= \frac{3}{4} = \%75$$

Sonuç 4: $güven(makarna \rightarrow ekmek, peynir) = \frac{sayı(Ekmek, makarna, peynir)}{sayı(makarna)}$

$$= \frac{3}{4} = \%75$$

Örnek 1.'e ait Birliktelik Kuralları

Birliktelik kuralı	Anlamı	Güven
Ekmek&Makarna→Peynir	Ekmek ve Makarnanın bulunduğu ürün kümesinde Peynirin olma olasılığı	%100
Ekmek→Peynir&Makarna	Ekmeğin yer aldığı bir ürün kümesinde peynir ve makarnanın olma olasılığı	%75
Peynir→Ekmek&Makarna	Peynirin yer aldığı bir ürün kümesinde ekmek ve makarnanın olma olasılığı	%75
Makarna→Ekmek&Peynir	Makarnanın yer aldığı bir ürün kümesinde ekmek ve peynirin olma olasılığı	%75