

# データをあつめるのが 苦手なフレンズ



協調フィルタリングしたい.....したくない？

	商品1	商品2	商品3	商品4	商品5	商品6	商品7	商品8	商品9	商品10	顧客Xとの相関関係
顧客X	-	1	0	-	-	-	-	0	0	1	1.000
顧客A	1	1	1	-	-	-	-	0	0	0	0.167
顧客B	-	-	-	0	0	0	1	1	1	0	-1.000
顧客C	0	1	0	0	-	1	1	0	0	1	1.000
顧客D	0	-	-	0	1	1	0	0	1	1	0.500
顧客E	-	1	0	-	1	0	-	0	0	0	0.612
おすすめ度	0.00			0.00	1.00	0.67	0.50				

ユーザって言われても...  
うち実店舗ですし...  
ECじゃないですし...







あっ...あの！ POS のデータが使えませんか！？

なにそれー！？　すごーい！！



## POS (POS レジ)とは何?

### POSレジでフレッシュなお店づくりを実現

POSとは (Point Of Sale) の略で、直訳しますと、販売時点という意味ですがPOSシステムやPOSレジを省略してPOSと呼んでいることとなります。



つまり、POSレジとは売上が発生する時点 (お客様が商品をお買い上げになる時点) で、その買上げ商品の値札に付与されているバーコード情報をスキャナーで (読み取り機) で瞬時に読み取り、その商品の部門、品名、値段などを画面に表示しシートに印刷します。

それと同時にレジ本体の記憶部 (メモリー) に各種情報を記録します。その記録されたデータを集計分析し、売れ筋や死に筋商品を早く正確に見つけ出し、商品を一日でも早く補充し、不良在庫商品を値下げなどで処分し、店舗の在庫の回転を計るのが売上商品管理です。

また、いつ、どこのお客様がどんな商品を買っていただいたかなどを記憶させ、集計することによりお客様に対するサービスの向上や、売上データを分析して売上の確保や顧客満足度を上げることができます。

# POS (Point Of Sale)

商品がレジを通った記録が取れる

これで.....分析ができるね!

## アソシエーション分析(1)

### 1. アソシエーション分析とは

POS(Point Of Sales : 販売時点情報管理)データが獲得しやすくなった昨今、重要な課題の1つは蓄積されたデータを有効に活用することである。説明の便利のため、POSデータをイメージした架空のデータを表1に示す。通常このようなデータをマーケット・バスケット・トランザクション(market basket transaction)と呼ぶ。表1の各行をそれぞれ1つのトランザクション(transaction,取引)、あるいはバスケットと呼ぶ。

表1 買い物バスケットの事例  
(TID: Transaction ID)

TID	アイテム集合
1	{パン、牛乳、ハム、果物}
2	{パン、オムツ、ビール、ハム}
3	{ソーセージ、ビール、オムツ}
4	{弁当、ビール、オムツ、タバコ}
5	{弁当、ビール、オレンジジュース、果物}



表1 買い物バスケットの事例

(TID: Transaction ID)

TID	アイテム集合
1	{パン、牛乳、ハム、果物}
2	{パン、オムツ、ビール、ハム}
3	{ソーセージ、ビール、オムツ}
4	{弁当、ビール、オムツ、タバコ}
5	{弁当、ビール、オレンジジュース、果物}

$$\text{supp}(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{M}$$

$$\text{conf}(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} = \frac{\text{supp}(X \Rightarrow Y)}{\text{supp}(X)}$$

```
> data.ap1<-apriori(data.tran,parameter = list(supp =0.2,maxlen=3))
```

```
> inspect(head(SORT(data.ap1, by = "support"), n = 20))
```

	lhs	rhs	support	confidence	lift
1	{}	=> {ビール}	0.8	0.8	1.000000
2	{オムツ}	=> {ビール}	0.6	1.0	1.250000
3	{ハム}	=> {パン}	0.4	1.0	2.500000
4	{パン}	=> {ハム}	0.4	1.0	2.500000
5	{弁当}	=> {ビール}	0.4	1.0	1.250000
6	{ソーセージ}	=> {オムツ}	0.2	1.0	1.666667
7	{ソーセージ}	=> {ビール}	0.2	1.0	1.250000
8	{牛乳}	=> {ハム}	0.2	1.0	2.500000
9	{牛乳}	=> {果物}	0.2	1.0	2.500000
10	{牛乳}	=> {パン}	0.2	1.0	2.500000
11	{タバコ}	=> {弁当}	0.2	1.0	2.500000
12	{タバコ}	=> {オムツ}	0.2	1.0	1.666667
13	{タバコ}	=> {ビール}	0.2	1.0	1.250000
14	{オレンジジュース}	=> {果物}	0.2	1.0	2.500000
15	{オレンジジュース}	=> {弁当}	0.2	1.0	2.500000
16	{オレンジジュース}	=> {ビール}	0.2	1.0	1.250000
17	{オムツ, ソーセージ}	=> {ビール}	0.2	1.0	1.250000
18	{ビール, ソーセージ}	=> {オムツ}	0.2	1.0	1.666667
19	{牛乳, ハム}	=> {果物}	0.2	1.0	2.500000
20	{牛乳, 果物}	=> {ハム}	0.2	1.0	2.500000

$$\text{lift}(X \Rightarrow Y) = \frac{\text{conf}(X \Rightarrow Y)}{\text{supp}(Y)}$$



表1 買い物バスケットの事例  
(TID: Transaction ID)

TID	アイテム集合
1	{パン、牛乳、ハム、果物}
2	{パン、オムツ、ビール、ハム}
3	{ソーセージ、ビール、オムツ}
4	{弁当、ビール、オムツ、タバコ}
5	{弁当、ビール、オレンジジュース、果物}

```
> data.ap1<-apriori(data.tran,parameter = list(supp =0.2,maxlen=3))
```

```
> inspect(head(SORT(data.ap1, by = "support"), n = 20))
```

	lhs	rhs	support	confidence	lift
1	{}	=> {ビール}	0.8	0.8	1.000000
2	{オムツ}	=> {ビール}	0.6	1.0	1.250000
3	{ハム}	=> {パン}	0.4	1.0	2.500000
4	{パン}	=> {ハム}	0.4	1.0	2.500000
5	{弁当}	=> {ビール}	0.4	1.0	1.250000
6	{ソーセージ}	=> {オムツ}	0.2	1.0	1.666667
7	{ソーセージ}	=> {ビール}	0.2	1.0	1.250000
8	{牛乳}	=> {ハム}	0.2	1.0	2.500000
9	{牛乳}	=> {果物}	0.2	1.0	2.500000
10	{牛乳}	=> {パン}	0.2	1.0	2.500000
11	{タバコ}	=> {弁当}	0.2	1.0	2.500000
12	{タバコ}	=> {オムツ}	0.2	1.0	1.666667
13	{タバコ}	=> {ビール}	0.2	1.0	1.250000
14	{オレンジジュース}	=> {果物}	0.2	1.0	2.500000
15	{オレンジジュース}	=> {弁当}	0.2	1.0	2.500000
16	{オレンジジュース}	=> {ビール}	0.2	1.0	1.250000
17	{オムツ, ソーセージ}	=> {ビール}	0.2	1.0	1.250000
18	{ビール, ソーセージ}	=> {オムツ}	0.2	1.0	1.666667
19	{牛乳, ハム}	=> {果物}	0.2	1.0	2.500000
20	{牛乳, 果物}	=> {ハム}	0.2	1.0	2.500000

$$supp(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{M}$$

$$conf(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} = \frac{supp(X \Rightarrow Y)}{supp(X)}$$

$$lift(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{supp(Y)}$$

$$Support(X \rightarrow Y) = p(X, Y)$$

$$Confidence(X \rightarrow Y) = p(Y|X) = \frac{p(X, Y)}{p(X)}$$

$$Lift(X \rightarrow Y) = \frac{p(Y|X)}{p(Y)} = \frac{p(X, Y)}{p(X)p(Y)}$$

表1 買い物バスケットの事例  
(TID: Transaction ID)

TID	アイテム集合
1	{パン、牛乳、ハム、果物}
2	{パン、オムツ、ビール、ハム}
3	{ソーセージ、ビール、オムツ}
4	{弁当、ビール、オムツ、タバコ}
5	{弁当、ビール、オレンジジュース、果物}

```
> data.ap1<-apriori(data.tran,parameter = list(supp =0.2,maxlen=3))
> inspect(head(SORT(data.ap1, by = "support"), n = 20))
```

	lhs	rhs	support	confidence	lift
1	{}	=> {ビール}	0.8	0.8	1.000000
2	{オムツ}	=> {ビール}	0.6	1.0	1.250000
3	{ハム}	=> {パン}	0.4	1.0	2.500000
4	{パン}	=> {ハム}	0.4	1.0	2.500000
5	{弁当}	=> {ビール}	0.4	1.0	1.250000
6	{ソーセージ}	=> {オムツ}	0.2	1.0	1.888887
7	{ソーセージ}	=> {ビール}	0.2	1.0	1.250000
8	{牛乳}	=> {ハム}	0.2	1.0	2.500000
9	{牛乳}	=> {果物}	0.2	1.0	2.500000
10	{牛乳}	=> {パン}	0.2	1.0	2.500000
11	{タバコ}	=> {弁当}	0.2	1.0	2.500000
12	{タバコ}	=> {オムツ}	0.2	1.0	1.888887
13	{タバコ}	=> {ビール}	0.2	1.0	1.250000
14	{オレンジジュース}	=> {果物}	0.2	1.0	2.500000
15	{オレンジジュース}	=> {弁当}	0.2	1.0	2.500000
16	{オレンジジュース}	=> {ビール}	0.2	1.0	1.250000
17	{オムツ, ソーセージ}	=> {ビール}	0.2	1.0	1.250000
18	{ビール, ソーセージ}	=> {オムツ}	0.2	1.0	1.888887
19	{牛乳, ハム}	=> {果物}	0.2	1.0	2.500000
20	{牛乳, 果物}	=> {ハム}	0.2	1.0	2.500000

$$Support(X \rightarrow Y) = p(X, Y)$$

XとYが共起する確率

$$Confidence(X \rightarrow Y) = p(Y|X) = \frac{p(X, Y)}{p(X)}$$

Xを買ったとき、Yも買う確率

Xが出てくる母集団中のYが出てくる確率

$$Lift(X \rightarrow Y) = \frac{p(Y|X)}{p(Y)} = \frac{p(X, Y)}{p(X)p(Y)}$$

全体に対するYが出てくる確率

対

Xが出てくる母集団中のYが出てくる確率

# 雑談するスペース

Askさんがリツイートしました



ギズモード・ジャパン @gizmodojapan

1h

英国離脱とトランプ当選、実はある会社に関係していた？

[gizmodo.jp/2017/02/trump-...](https://gizmodo.jp/2017/02/trump-...)



42



54