Frequent Patterns and Association Rules

[redacted]

PART 01

Frequent Pattern Analysis

1

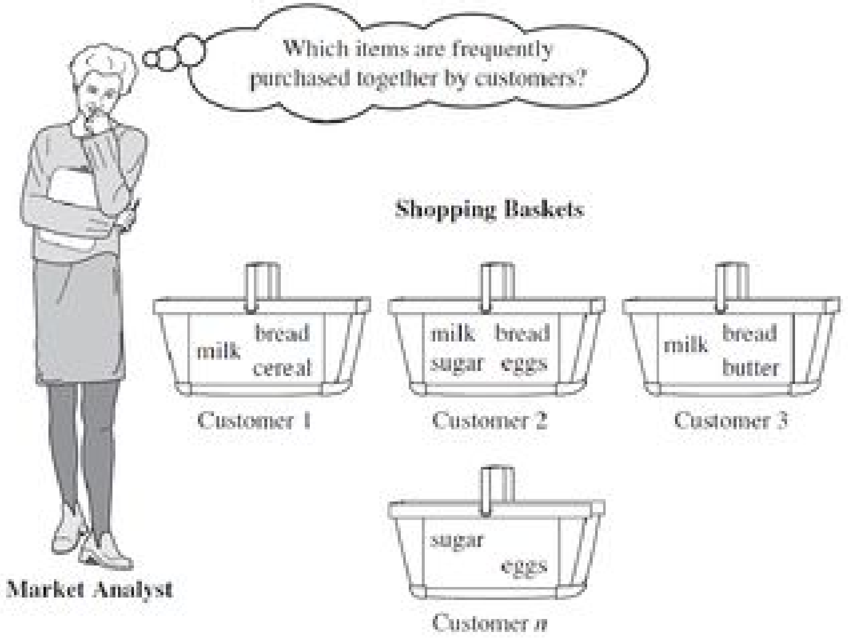
Frequent Pattern Analysis

- Frequent pattern: a pattern (a set of items, subsequences, substructures, etc.) that occurs frequently in a data set
- Motivation: Finding inherent regularities in data
  - What products were often purchased together?
  - What are the subsequent purchases after buying a PC?
  - What kinds of DNA are sensitive to this new drug?
  - Can we automatically classify web documents?

1

What Is Frequent Pattern Analysis?

- Applications
  - Basket data analysis
  - cross-marketing
  - catalog design
  - sale campaign analysis
  - Web log (click stream) analysis
  - DNA sequence analysis



1

Why Is Freq. Pattern Mining Important?

- An intrinsic and important property of datasets

1 Foundation for many essential data mining tasks

- Association, correlation, and causality analysis
- Sequential, structural (e.g., sub-graph) patterns
- Pattern analysis in spatiotemporal, multimedia, time-series, and stream data
- Classification: discriminative, frequent pattern analysis
- Cluster analysis: frequent pattern-based clustering

1

Mining Association Rules

Transaction-id	Items bought
1	A, B, D
2	A, C, D
3	A, D, E
4	B, E, F
5	B, C, D, E, F

- Transaction data analysis. Given:
  - A database of transactions (Each tx. has a list of items purchased)
  - Minimum confidence and minimum support
- Find all association rules: the presence of one set of items implies the presence of another set of items

Diaper → Beer [0.5%, 75%]
(support, confidence)

1 Mining Strong Association Rules in Transaction Databases (1/2)

- Measurement of rule strength in a transaction database.

A→B [*support*, *confidence*]

$$\text{support}(A \cup B) = \Pr(A \cup B) = \frac{\#\text{of tx containing all items in } A \cup B}{\text{total}\#\text{ of tx}}$$

$$\text{confidence}(A \cup B) = \Pr(B \mid A) = \frac{\#\text{of tx containing both } A \cup B}{\#\text{ of tx containing A}}$$

Note: support(AUB) means support for occurrences of transactions **X and Y both appear**. “U” is NOT the logical OR here.

1

Example of Association Rules

Transaction-id	Items bought
1	A, B, D
2	A, C, D
3	A, D, E
4	B, E, F
5	B, C, D, E, F

Let min. support = 50%, min. confidence = 50%

- Frequent patterns: {A:3, B:3, D:4, E:3, AD:3}
- Association rules: **A → D (s = 60%, c = 100%)**
  D → A (s = 60%, c = 75%)

1 Mining Strong Association Rules in Transaction Databases (2/2)

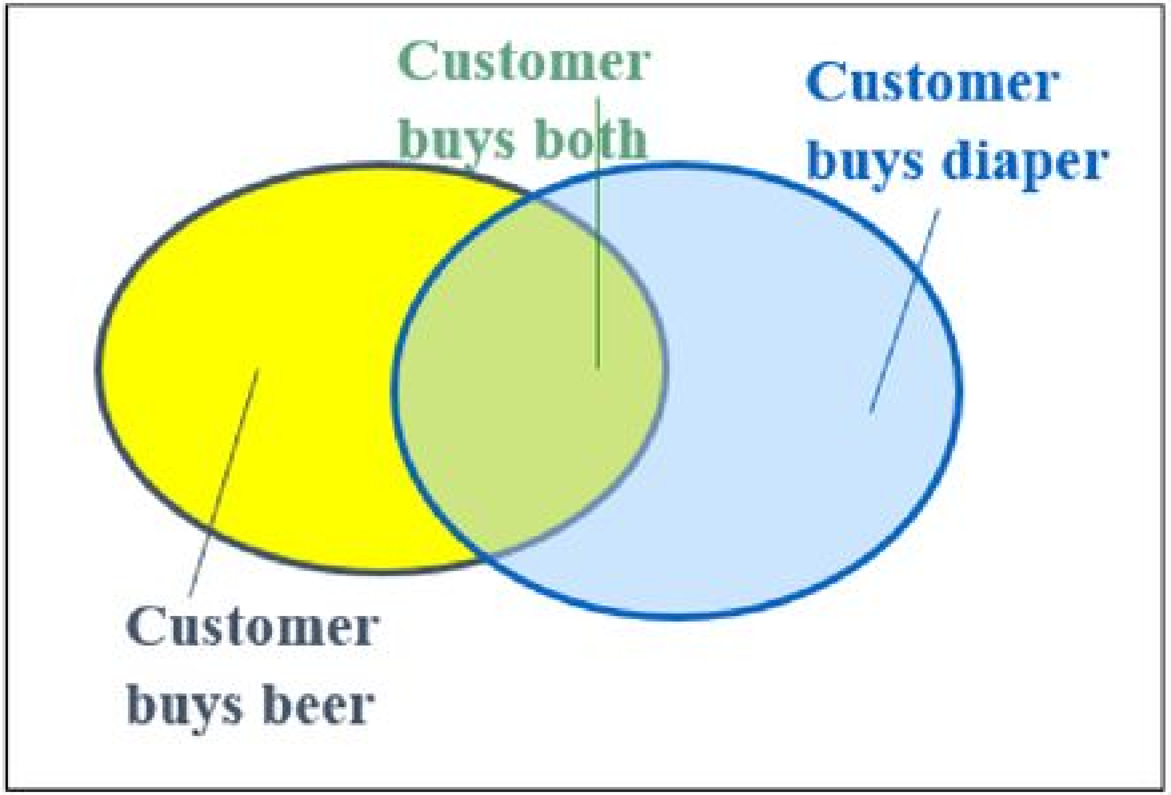
- We are often interested in only strong associations, i.e.,
  - ***support* $\geq$ *min_sup***
  - ***confidence* $\geq$ *min_conf***

- Examples:
  - milk $\rightarrow$ bread [5%, 60%]
  - tire and auto_accessories $\rightarrow$ auto_services [2%, 80%].

1

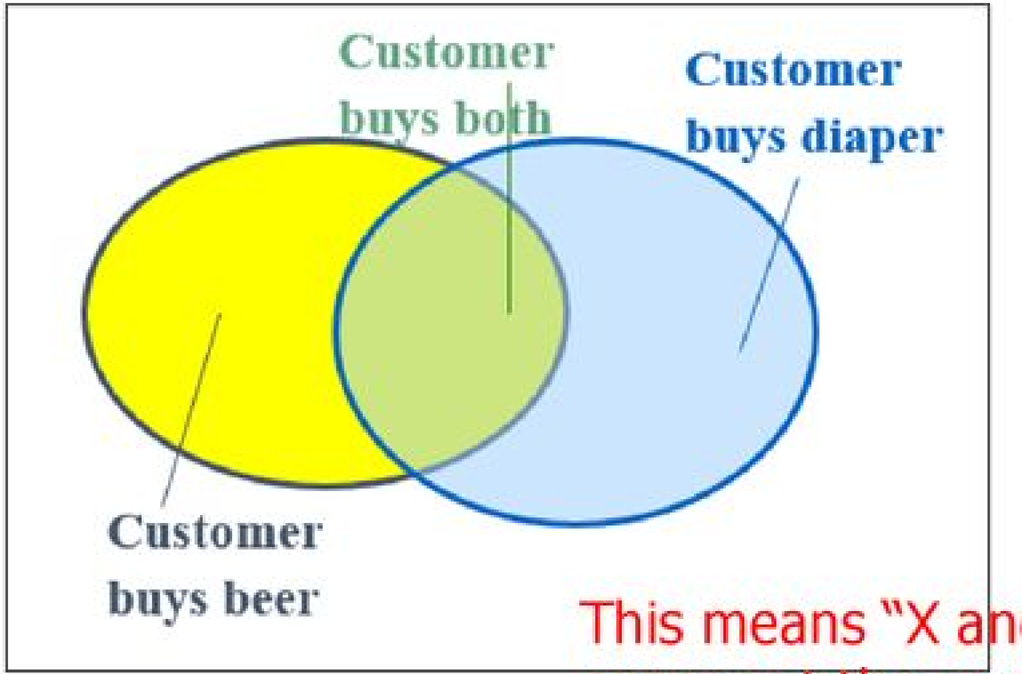
Figure to analyze

Tid	Items bought
10	Beer, Nuts, Diaper
20	Beer, Coffee, Diaper
30	Beer, Diaper, Eggs
40	Nuts, Eggs, Milk
50	Nuts, Coffee, Diaper, Eggs, Milk



1

Tid	Items bought
10	Beer, Nuts, Diaper
20	Beer, Coffee, Diaper
30	Beer, Diaper, Eggs
40	Nuts, Eggs, Milk
50	Nuts, Coffee, Diaper, Eggs, Milk



This means "X and Y" appear at the same time

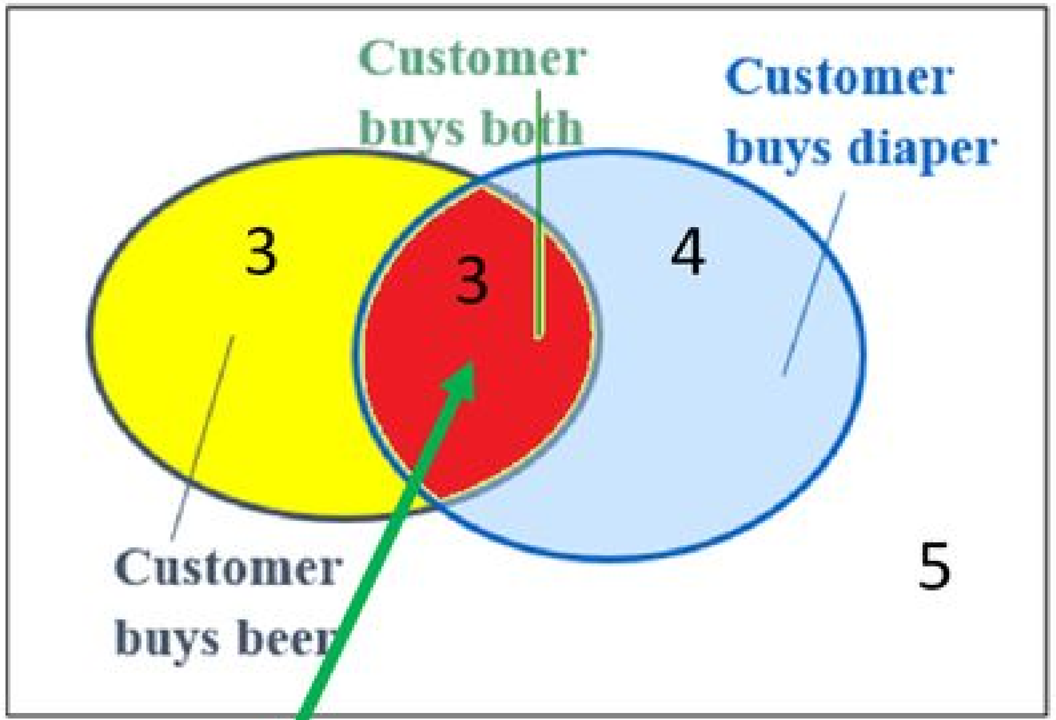
- **support**, *s*, probability that a transaction contains $X \cup Y$
- **confidence**, *c*, conditional probability that a transaction having X also contains *Y*

$$support(A \Rightarrow B) = P(A \cup B)$$
$$confidence(A \Rightarrow B) = P(B|A).$$

$$confidence(A \Rightarrow B) = P(B|A) = \frac{support(A \cup B)}{support(A)} = \frac{support\_count(A \cup B)}{support\_count(A)}$$

1

Tid	Items bought
10	Beer, Nuts, Diaper
20	Beer, Coffee, Diaper
30	Beer, Diaper, Eggs
40	Nuts, Eggs, Milk
50	Nuts, Coffee, Diaper, Eggs, Milk

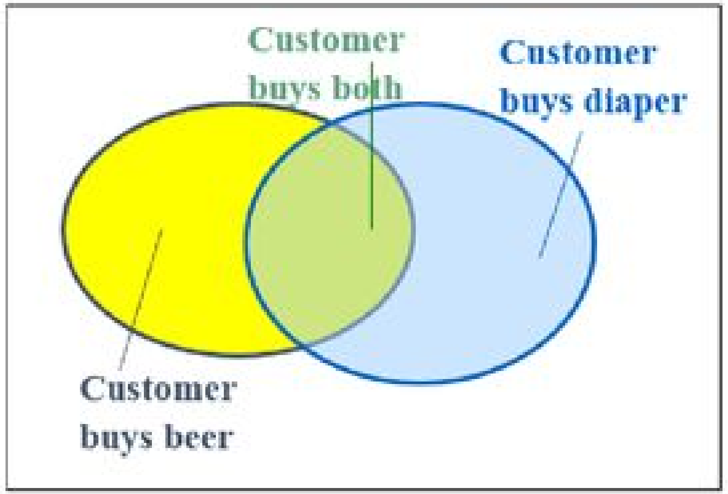


- support(Beer $\cup$ Diaper)=3/5=60%
- confidence(Beer=>Diaper)=3/3=100%

1

Basic Concepts: Frequent Patterns

Tid	Items bought
10	Beer, Nuts, Diaper
20	Beer, Coffee, Diaper
30	Beer, Diaper, Eggs
40	Nuts, Eggs, Milk
50	Nuts, Coffee, Diaper, Eggs, Milk



- Find all the rules $X \rightarrow Y$ with minimum support and confidence
- *Let minsup = 50%, minconf = 50%*
- *Frequent Pattern:*
  Beer:3, Nuts:3, Diaper:4, Eggs:3, {Beer, Diaper}:3, ~~{Nuts, Diaper}:2~~
- *Association rules: (many more!)*
  - *Beer → Diaper (60%, 100%)*
  - *Diaper → Beer (60%, 75%)*

1 Two Steps for Mining Association Rules

- Determining "large (frequent) itemsets"
  - The main factor for overall performance
  - The **downward closure property** of frequent patterns
    - Any subset of a frequent itemset must be frequent
    - If {beer, diaper, nuts} is frequent, so is {beer, diaper}
    - i.e., every transaction having {beer, diaper, nuts} also contains {beer, diaper}
- Generating rules

Scalable Frequent Itemset Mining Methods

Apriori

A Candidate Generation-and-Test Approach

1

Brute-Force Approach

- List all items

TID	*List of item_IDs*
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I2
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

1

Brute-Force Approach

- List all items
- But…
  - 1) Unnecessary itemset examination, i.e., constructing too many non-frequent itemsets

TID	*List of item_IDs*
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I2
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

1

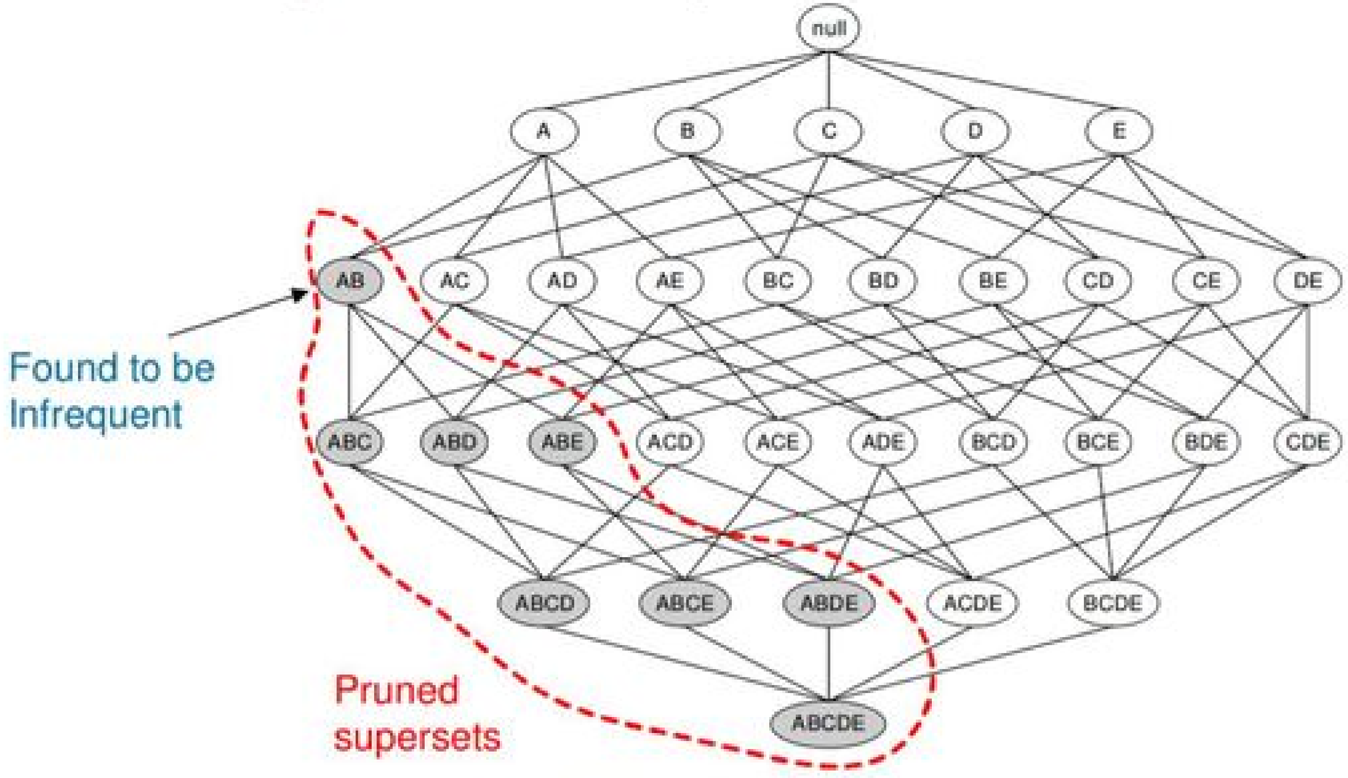
Brute-Force Approach

- List all items
- But…
  - 1) Unnecessary itemset examination, i.e., constructing too many non-frequent itemsets
  - 2) Generating duplicate itemsets

TID	*List of item_IDs*
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I2
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

To Avoid Unnecessary Examination

- Apriori Pruning Principle, a.k.a Downward Closure Property
- If there is any itemset which is infrequent, its superset should not be generated/tested!



Data Mining: Association Rules

Item	Count
Bread	4
Coke	2
Milk	4
Beer	3
Diaper	4
Eggs	1

Items (1-itemsets)

Itemset	Count
{Bread,Milk}	3
{Bread,Beer}	2
{Bread,Diaper}	3
{Milk,Beer}	2
{Milk,Diaper}	3
{Beer,Diaper}	3

Pairs (2-itemsets)

(No need to generate candidates involving Coke or Eggs)

Minimum Support = 3

Triplets (3-itemsets)

Itemset	Count
{Bread,Milk,Diaper}	3

If every subset is considered,

$$^6C_1 + {}^6C_2 + {}^6C_3 = 41$$

With support-based pruning,

$$6 + 6 + 1 = 13$$

1

To Avoid Generating Duplicate Itemsets

Join

- Joining 2 k-itemsets:
  2 k-itemsets I1, I2 could be joined into a (k+1)-itemset if I1 and I2 have the same first (k-1) items
- E.g., {A, B, C, D, E} join {A, B, C, D, F} = {A, B, C, D, E, F}
- However, {A, B, C, D, E} cannot join {A, B, C, E, F}

- Start with all 1-itemsets (C_1)
- Go through data and count their support and find all "large" 1-itemsets (L_1)
- Join them to form "candidate" 2-itemsets (C_2)
- Go through data and count their support and find all "large" 2-itemsets (L_2)
- Join them to form "candidate" 3-itemsets … (C_3)

…

C_1

Itemset	sup
{A}	2
{B}	3
{C}	3
{D}	1
{E}	3

→

L_1

Itemset	sup
{A}	2
{B}	3
{C}	3
{E}	3

→

C_2

Itemset	sup
{A, B}	1
{A, C}	2
{A, E}	1
{B, C}	2
{B, E}	3
{C, E}	2

…

- Large itemset (L_k): itemset with support > s
- Candidate itemset (C_k): itemset that may have support > s

min. support =2 tx's (50%)

The Apriori Algorithm—An Example

Database TDB

Tid	Items
100	A, C, D
200	B, C, E
300	A, B, C, E
400	B, E

1st scan →

C_1

Itemset	sup
{A}	2
{B}	3
{C}	3
{D}	1
{E}	3

L_1

Itemset	sup
{A}	2
{B}	3
{C}	3
{E}	3

C_2

Itemset
{A, B}
{A, C}
{A, E}
{B, C}
{B, E}
{C, E}

2nd scan ←

C_2

Itemset	sup
{A, B}	1
{A, C}	2
{A, E}	1
{B, C}	2
{B, E}	3
{C, E}	2

L_2

Itemset	sup
{A, C}	2
{B, C}	2
{B, E}	3
{C, E}	2

C_3

Itemset
{B, C, E}

3rd scan →

L_3

Itemset	sup
{B, C, E}	2

min. support =2 tx's (50%)

The Apriori Algorithm—An Example

Database TDB

Tid	Items
100	A, C, D
200	B, C, E
300	A, B, C, E
400	B, E

1st scan →

C_1

Itemset	sup
{A}	2
{B}	3
{C}	3
{D}	1
{E}	3

→

L_1

Itemset	sup
{A}	2
{B}	3
{C}	3
{E}	3

C_2

Item
{A,
{A,
{A,
{B,
{B,
{C,

←

L_2

Itemset	sup
{A, C}	2
{B, C}	2
{B, E}	3
{C, E}	2

If {A,B,C} is a frequent item set, then {A,C}, {A,B} must exist in L2 as well.

Here, {A,B} does not exist in L2, so no bother checking!

C_3

Itemset
{B, C, E}

3rd scan →

L_3

Itemset	sup
{B, C, E}	2

From Large Itemsets to Rules

1

Database TDB

Tid	Items
100	A, C, D
200	B, C, E
300	A, B, C, E
400	B, E

- Recall that confidence is defined as:

$$confidence(A \Rightarrow B) = P(B|A) = \frac{support\_count(A \cup B)}{support\_count(A)}$$

- For each large itemset *l*
  - For each subset *s* of *l*, if (*sup(l) / sup(s) ≥ min_conf*)
    - output the rule ***s=> (l-s)***
      - conf. = sup(l)/sup(s)
      - support = sup(l)

L_2

Itemset	sup
{A, C}	2
{B, C}	2
{B, E}	3
{C, E}	2

L_3

Itemset	sup
{B, C, E}	2

- E.g., *l*={B, C, E} with support =2 is a frequent 3-item set, assume min_conf=80%
- Let *s*={C,E}, *sup(l) / sup(s)* =2/2=100%>80%
- Therefore, {C,E}=>{B} is an association rule with support = 50%, confidence = 100%

1

Redundant Rules

- For the same support and confidence, if we have a rule {a,d} →{c,e,f,g}, do we have:
  - {a,d} →{c,e,f} ? Yes!
  - {a} →{c,e,f,g} ? No!
    - **Consider the example in previous page**
    - *l*={B, C, E}, *s*={C,E}, then {C,E}->{B} is an association rule
    - **However,** *l*={B, C, E}, *s*={C}, i.e., {C}->{B} is not
  - {a,d,c} →{e,f,g} ? No!
  - {a} →{c,d,e,f,g} ? No!

PART 02

Which Patterns Are Interesting? Pattern Evaluation Methods

2 Strong rules are not necessarily interesting

- 10,000 transactions:
  6,000 include games, 7,500 include videos, 4,000 include both game and video
- A **strong** association rule is thus derived (min_sup=30%, min_conf=60%):

$$buys(X, \text{"computer games"}) \Rightarrow buys(X, \text{"videos"})$$

$$[support = 40\%, confidence = 66\%].$$

- This rule is **MISLEADING**, because probability of videos is 75% > 66% (buying game and video together)
- In fact, games and videos are negatively associated
  Buying one actually decreases the likelihood of buying the other

2 Interestingness Measure: Correlations (Lift)

- *games* $\Rightarrow$ ***not*** *video* [20%, 33.3%] is more accurate, although with lower support and confidence
- Measure of dependent/correlated events: lift

  If the occurrence of A is independent of B => $P(A \cup B) = P(A)P(B)$

$$\mathrm{lift(A,B)} = \frac{P(A \cup B)}{P(A)P(B)}$$

$\mathrm{lift(A,B)} < 1$ => A, B are negative correlated

$\mathrm{lift(A,B)} > 1$ => A, B are positively correlated

$\mathrm{lift(A,B)} = 1$ => A, B are independent

$$lift(G,V) = \frac{4000/10000}{6000/10000 * 7500/10000} = 0.89$$

$$lift(G,\neg V) = \frac{2000/10000}{6000/10000 * 2500/10000} = 1.33$$

	Game	Not game	Sum (row)
Video	4000	3500	7500
Not video	2000	500	2500
Sum(col.)	6000	4000	10000

2

Interesting Measure: Chi-square(X^2)

	game	$\overline{game}$	Σ_{row}
video	4000 (4500)	3500 (3000)	7500
$\overline{video}$	2000 (1500)	500 (1000)	2500
Σ_{col}	6000	4000	10,000

$$\chi^2 = \Sigma \frac{(observed - expected)^2}{expected}$$

$$= \frac{(4000-4500)^2}{4500} + \frac{(3500-3000)^2}{3000} + \frac{(2000-1500)^2}{1500} + \frac{(500-1000)^2}{1000}$$

$$= 555.6.$$

χ^2>1, and the observed value of *(game, video)=4000*, which is less than the expected value *4,5000*

$\Rightarrow$ Buying game and buying video are negatively correlated

Consistent with the conclusion derived from the analysis of the lift measure

2

Pattern Evaluation Measures

- **all_confidence:** Minimum confidence of “A=>B” and “B=>A”

$$all\_conf(A,B) = \frac{sup(A \cup B)}{max\{sup(A), sup(B)\}} = min\{P(A|B), P(B|A)\}$$

- **max_confidence:** Maximum confidence of “A=>B” and “B=>A”

$$max\_conf(A, B) = max\{P(A \mid B), P(B \mid A)\}$$

- **Kulczynski:** Average confidence of “A=>B” and “B=>A”

$$Kulc(A, B) = \frac{1}{2}(P(A|B) + P(B|A))$$

- **Cosine:** $$cosine(A, B) = \frac{P(A \cup B)}{\sqrt{P(A) \times P(B)}} = \frac{sup(A \cup B)}{\sqrt{sup(A) \times sup(B)}} = \sqrt{P(A|B) \times P(B|A)}.$$

2

Comparison of Interestingness Measures

m and **c**:

- positively correlated in D_1, D_2, i.e., mc(10,000) > $\overline{m}c$(1,000)=$m\overline{c}$(1,000)
- negatively correlated in D_3
- neutral in D_4

	milk	$\overline{milk}$	Σ_{row}
coffee	mc	$\overline{m}c$	c
$\overline{coffee}$	$m\overline{c}$	$\overline{mc}$	$\overline{c}$
Σ_{col}	m	$\overline{m}$	Σ

Data Set	mc	$\overline{m}c$	$m\overline{c}$	$\overline{mc}$	χ^2	*lift*	*all_conf.*	*max_conf.*	*Kulc.*	*cosine*
D_1	10,000	1000	1000	100,000	90557	9.26	0.91	0.91	0.91	0.91
D_2	10,000	1000	1000	100	0	1	0.91	0.91	0.91	0.91
D_3	100	1000	1000	100,000	670	8.44	0.09	0.09	0.09	0.09
D_4	1000	1000	1000	100,000	24740	25.75	0.5	0.5	0.5	0.5
D_5	1000	100	10,000	100,000	8173	9.18	0.09	0.91	0.5	0.29
D_6	1000	10	100,000	100,000	965	1.97	0.01	0.99	0.5	0.10

2 Comparison of Interestingness Measures

m and **c**:

- positively correlated in D_1, D_2,
  i.e., $mc(10{,}000) > \overline{m}c(1{,}000)=m\bar{c}(1{,}000)$
- negatively correlated in D_3.
- neutral in D_4

All the four new measures show m and c are strongly positively associated

Data Set	mc	$\overline{m}c$	$m\bar{c}$	$\overline{mc}$	χ^2	*lift*	*all_conf.*	*max_conf.*	*Kulc.*	*cosine*
D_1	10,000	1000	1000	100,000	90557	9.26	0.91	0.91	0.91	0.91
D_2	10,000	1000	1000	100	0	1	0.91	0.91	0.91	0.91
D_3	100	1000	1000	100,000	670	8.44	0.09	0.09	0.09	0.09
D_4	1000	1000	1000	100,000	24740	25.75	0.5	0.5	0.5	0.5
D_5	1000	100	10,000	100,000	8173	9.18	0.09	0.91	0.5	0.29
D_6	1000	10	100,000	100,000	965	1.97	0.01	0.99	0.5	0.10

2

Comparison of Interestingness Measures

m and **c**:

- positively correlated in D_1, D_2,
- negatively correlated in D_3
- neutral in D_4

In real-world scenarios, $\overline{mc}$ is usually huge and unstable

χ^2 and lift generate dramatically different measures
Due to their sensitivity to $\overline{mc}$

Data Set	mc	$\overline{m}c$	$m\overline{c}$	$\overline{mc}$	χ^2	*lift*	*all_conf.*	*max_conf.*	*Kulc.*	*cosine*
D_1	10,000	1000	1000	100,000	90557	9.26	0.91	0.91	0.91	0.91
D_2	10,000	1000	1000	100	0	1	0.91	0.91	0.91	0.91
D_3	100	1000	1000	100,000	670	8.44	0.09	0.09	0.09	0.09
D_4	1000	1000	1000	100,000	24740	25.75	0.5	0.5	0.5	0.5
D_5	1000	100	10,000	100,000	8173	9.18	0.09	0.91	0.5	0.29
D_6	1000	10	100,000	100,000	965	1.97	0.01	0.99	0.5	0.10

2 Comparison of Interestingness Measures

m and **c**:

- positively correlated in D_1, D_2,
- negatively correlated in D_3
- neutral in D_4

All the four new measures show m and c are strongly nagtively associated

Data Set	mc	$\overline{m}c$	$m\overline{c}$	$\overline{mc}$	χ^2	*lift*	*all_conf.*	*max_co*	*Kulc.*	*cosine*
D_1	10,000	1000	1000	100,000	90557	9.26	0.91	0.91	0.91	0.91
D_2	10,000	1000	1000	100	0	1	0.91	0.91	0.91	0.91
D_3	100	1000	1000	100,000	670	8.44	0.09	0.09	0.09	0.09
D_4	1000	1000	1000	100,000	24740	25.75	0.5	0.5	0.5	0.5
D_5	1000	100	10,000	100,000	8173	9.18	0.09	0.91	0.5	0.29
D_6	1000	10	100,000	100,000	965	1.97	0.01	0.99	0.5	0.10

2 Comparison of Interestingness Measures

m and **c**:

- positively correlated in D_1, D_2,
- negatively correlated in D_3
- neutral in D_4

χ^2 and lift: values are between D_1 and D_2

Data Set	mc	$\overline{m}c$	$m\overline{c}$	$\overline{mc}$	χ^2	li[illegible]	all_conf.	max_conf.	Kulc.	cosine
D_1	10,000	1000	1000	100,000	90557	[illegible].26	0.91	0.91	0.91	0.91
D_2	10,000	1000	1000	100	0	1	0.91	0.91	0.91	0.91
D_3	100	1000	1000	100,000	670	8.44	0.09	0.09	0.09	0.09
D_4	1000	1000	1000	100,000	24740	25.75	0.5	0.5	0.5	0.5
D_5	1000	100	10,000	100,000	8173	9.18	0.09	0.91	0.5	0.29
D_6	1000	10	100,000	100,000	965	1.97	0.01	0.99	0.5	0.10

2 Comparison of Interestingness Measures

m and **c**:

- positively correlated in D_1, D_2,
- negatively correlated in D_3
- neutral in D_4

χ^2 and lift: show that D4 is positive associated between m and c

Data Set	mc	$\overline{m}c$	$m\overline{c}$	$\overline{mc}$	χ^2	lift	_conf.	max_conf.	Kulc.	cosine
D_1	10,000	1000	1000	100,000	90557	9.26	.91	0.91	0.91	0.91
D_2	10,000	1000	1000	100	0	1	0.91	0.91	0.91	0.91
D_3	100	1000	1000	100,000	670	8.44	0.09	0.09	0.09	0.09
D_4	1000	1000	1000	100,000	24740	25.75	0.5	0.5	0.5	0.5
D_5	1000	100	10,000	100,000	8173	9.18	0.09	0.91	0.5	0.29
D_6	1000	10	100,000	100,000	965	1.97	0.01	0.99	0.5	0.10

Comparison of Interestingness Measures

m and **c**:

- positively correlated in D_1, D_2,
- negatively correlated in D_3
- neutral in D_4

It is neutral as indicated by the four measures.
A customer buys coffee (or milk), the probability of buying milk (of coffee) is exactly 50%

Data Set	mc	$\overline{m}c$	$m\overline{c}$	$\overline{mc}$	χ^2	*lift*	*all_conf.*	*max_conf.*	*Kulc.*	*sine*
D_1	10,000	1000	1000	100,000	90557	9.26	0.91	0.91	0.91	91
D_2	10,000	1000	1000	100	0	1	0.91	0.91	0.91	.91
D_3	100	1000	1000	100,000	670	8.44	0.09	0.09	0.09	.09
D_4	1000	1000	1000	100,000	24740	25.75	0.5	0.5	0.5	0.5
D_5	1000	100	10,000	100,000	8173	9.18	0.09	0.91	0.5	0.29
D_6	1000	10	100,000	100,000	965	1.97	0.01	0.99	0.5	0.10

2

Why are lift X^2 and so poor?

- Null-transactions
  - Transaction that does not contain any of the itemsets being examined
  - E.g., $\overline{mc}$ is the number of null-transactions
- lift and χ^2 are strongly influenced by $\overline{mc}$
- The other four measures are good indicators
  - Their definitions remove the influence of $\overline{mc}$

2

Which Null-Invariant Measure Is Better?

- IR (Imbalance Ratio): measure the imbalance of two itemsets A and B in rule implications

$$IR(A,B) = \frac{|sup(A) - sup(B)|}{sup(A) + sup(B) - sup(A \cup B)}$$

- **c** occurs strongly suggests **m** occurs also
- **m** occurs strongly suggests **c** unlikely occur Diverse outcomes!!

Data Set	mc	$\overline{m}c$	$m\overline{c}$	$\overline{mc}$	χ^2	*lift*	*all_conf.*			
D_1	10,000	1000	1000	100,000	90557	9.26	0.91	0.91	0.91	0.91
D_2	10,000	1000	1000	100	0	1	0.91	0.91	0.91	0.91
D_3	100	1000	1000	100,000	670	8.44	0.09	0.09	0.09	0.09
D_4	1000	1000	1000	100,000	24740	25.75	0.5	0.5	0.5	0.5
D_5	1000	100	10,000	100,000	8173	9.18	0.09	0.91	0.5	0.29
D_6	1000	10	100,000	100,000	965	1.97	0.01	0.99	0.5	0.10

2

$$IR(A,B) = \frac{|sup(A) - sup(B)|}{sup(A) + sup(B) - sup(A \cup B)}$$

Which Null-Invariant Measure Is Better?

- Kulczynski and Imbalance Ratio (IR) together present a clear picture for all the three datasets D_4 through D_6
  - D_4 is balanced & neutral (IR(m,c)=0 => perfect balanced)
  - D_5 is imbalanced & neutral (IR(m,c)=0.89 => imbalanced)
  - D_6 is very imbalanced & neutral (IR(m,c)=0.99 => very skewed)

Data Set	mc	$\overline{m}c$	$m\overline{c}$	$\overline{mc}$	χ^2	*lift*	*all_conf.*	*max_conf.*	*Kulc.*	*cosine*
D_1	10,000	1000	1000	100,000	90557	9.26	0.91	0.91	0.91	0.91
D_2	10,000	1000	1000	100	0	1	0.91	0.91	0.91	0.91
D_3	100	1000	1000	100,000	670	8.44	0.09	0.09	0.09	0.09
D_4	1000	1000	1000	100,000	24740	25.75	0.5	0.5	0.5	0.5
D_5	1000	100	10,000	100,000	8173	9.18	0.09	0.91	0.5	0.29
D_6	1000	10	100,000	100,000	965	1.97	0.01	0.99	0.5	0.10

2017 THANK YOU!