RWTH AACHEN UNIVERSITY

Elements of Machine Learning & Data Science

Winter semester 2025/26

Lecture 6 – Association Rules

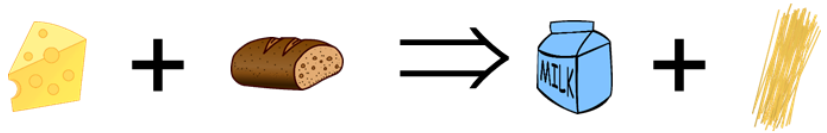
10.11.2025

Prof. Bastian Leibe

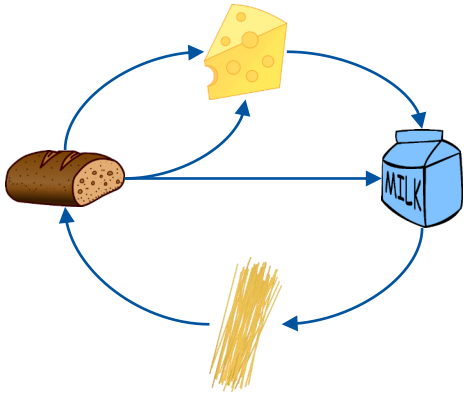
slides by Prof. Wil van der Aalst

Overview of the Lecture Topics

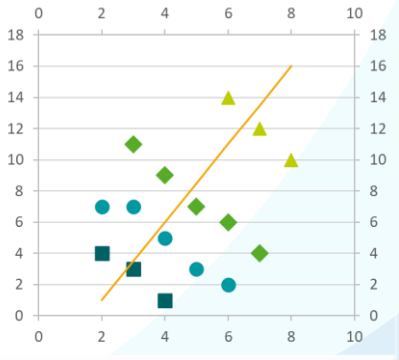
1. Introduction to Data Science
2. Decision Trees
3. Clustering
4. Frequent Itemsets
5. **Association Rules**
6. Time Series



Pattern Mining & Association Rules



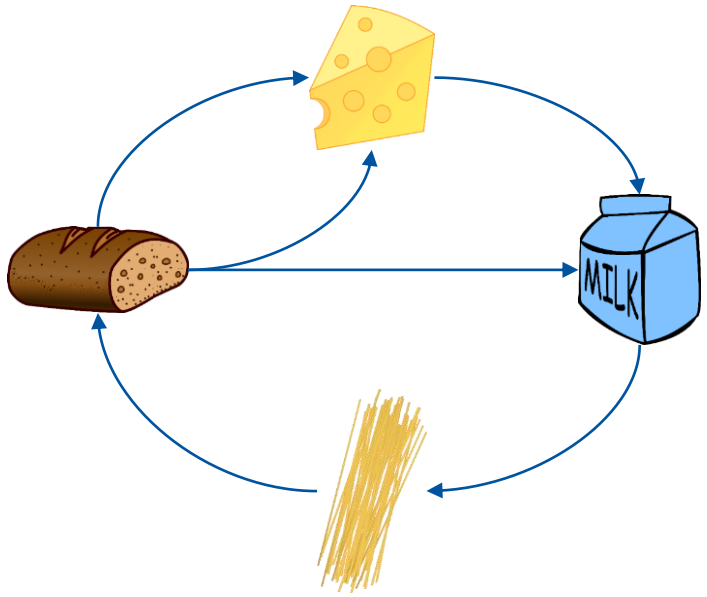
Generating Association Rules



Simpson's Paradox

Association Rules

1. **Introduction**
2. Generating Association Rules
3. Evaluation
4. Applications
5. Simpson’s Paradox



Association Rules - Notation

- $\mathcal{I} = \{I_1, I_2, \ldots, I_D\}$ is the set of all possible items
- A transaction $\mathcal{T} \in \mathbb{P}(\mathcal{I}) \setminus \{\emptyset\}$ is a non-empty itemset
- A dataset $\mathcal{X} \in \mathbb{M}(\mathbb{P}(\mathcal{I}))$ $(\text{such that } \emptyset \notin \mathcal{X})$ is a multiset of transactions
  (Here, $\mathbb{M}$ is the multiset and $\mathbb{P}$ is the powerset operator)

Same as before

- $\mathcal{A} \Rightarrow \mathcal{B}$ with $\mathcal{A} \subseteq \mathcal{I}, \mathcal{B} \subseteq \mathcal{I}$ and $\mathcal{A} \cap \mathcal{B} = \emptyset$ is an association rule
- For example, $\{\text{Cheese}, \text{Bread}\} \Rightarrow \{\text{Milk}\}$

$$\mathcal{A} \Rightarrow \mathcal{B}$$

Association Rules

{Cheese, Bread} ⇒ {Milk}
People that buy Cheese and Bread also tend to buy Milk.

{AC/DC, Queen} ⇒ {Metallica}
Users that listen to AC/DC and Queen also tend to listen to Metallica.

{Bitburger} ⇒ {Heineken, Palm}
People that buy Bitburger beer tend to buy both Heineken and Palm beer.

{Carbonara, Margherita } ⇒ {Espresso, Tiramisu}
People that buy Carbonara and Margherita also tend to buy Espresso and Tiramisu.

{part-245, part-345, part-456} ⇒ {part-372}
When Parts 245, 345, and 456 are replaced, then often also Part 372 is replaced.

From Frequent Itemsets to Association Rules

- Frequent Itemsets – a combinatorial explosion
- How to determine the interesting ones?
- How to turn itemsets into rules?

[1]

A Larger Supermarket May Have Up To 50000 Distinct Items

…and the number of rules exceeds the number of possible itemsets!

Support and Confidence

- Support: fraction of instances containing all items in $\mathcal{A} \cup \mathcal{B}$

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \text{support}(\mathcal{A} \cup \mathcal{B}) = \frac{\text{support\_count}(\mathcal{A} \cup \mathcal{B})}{\text{support\_count}(\emptyset)} = \frac{|[\mathcal{T} \in \mathcal{X} | \mathcal{A} \cup \mathcal{B} \subseteq \mathcal{T}]|}{|\mathcal{X}|}$$

Support and Confidence

- Support: fraction of instances containing all items in $\mathcal{A} \cup \mathcal{B}$

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \text{support}(\mathcal{A} \cup \mathcal{B}) = \frac{\text{support\_count}(\mathcal{A} \cup \mathcal{B})}{\text{support\_count}(\emptyset)} = \frac{|[\mathcal{T} \in \mathcal{X} | \mathcal{A} \cup \mathcal{B} \subseteq \mathcal{T}]|}{|\mathcal{X}|}$$

- Confidence: fraction of instances containing items in $\mathcal{A}$ which also contain items in $\mathcal{A} \cup \mathcal{B}$

$$\text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\text{support}(\mathcal{A} \cup \mathcal{B})}{\text{support}(\mathcal{A})} = \frac{\text{support\_count}(\mathcal{A} \cup \mathcal{B})}{\text{support\_count}(\mathcal{A})} = \frac{|[\mathcal{T} \in \mathcal{X} | \mathcal{A} \cup \mathcal{B} \subseteq \mathcal{T}]|}{|[\mathcal{T} \in \mathcal{X} | \mathcal{A} \subseteq \mathcal{T}]|}$$

Support and Confidence - Example

ID	Bought Items
1	{Bread, Cheese, Milk, Pasta}
2	{Bread, Cheese, Chips}
3	{Cheese, Pasta, Milk}
4	{Bread, Cheese, Milk}
5	{Bread, Pasta}

All three items Bread, Cheese and Milk need to be in the transaction to count

$$\text{support}(\{\text{Bread}\} \Rightarrow \{\text{Cheese}, \text{Milk}\}) = \text{support}(\{\text{Bread}, \text{Cheese}, \text{Milk}\}) = \frac{2}{5}$$

Support and Confidence - Example

ID	Bought Items
1	{Bread, Cheese, Milk, Pasta}
2	{Bread, Cheese, Chips}
3	{Cheese, Pasta, Milk}
4	{Bread, Cheese, Milk}
5	{Bread, Pasta}

$$\begin{aligned}\text{support}(\{\text{Bread}\} \Rightarrow \{\text{Cheese}, \text{Milk}\}) &= \text{support}(\{\text{Bread}, \text{Cheese}, \text{Milk}\}) = \tfrac{2}{5} \\ &= \text{support}(\{\text{Cheese}, \text{Milk}\} \Rightarrow \{\text{Bread}\}) \\ &= \text{support}(\{\text{Bread}, \text{Cheese}\} \Rightarrow \{\text{Milk}\})\end{aligned}$$

Symmetric: moving an item does not change the value

Support and Confidence - Example

ID	Bought Items
1	{Bread, Cheese, Milk, Pasta}
2	{Bread, Cheese, Chips}
3	{Cheese, Pasta, Milk}
4	{Bread, Cheese, Milk}
5	{Bread, Pasta}

$$\text{conf}(\{\text{Bread}\} \Rightarrow \{\text{Cheese}, \text{Milk}\}) = \frac{\text{support}(\{\text{Bread}, \text{Cheese}, \text{Milk}\})}{\text{support}(\{\text{Bread}\})} = \frac{2}{4}$$

Support and Confidence - Example

ID	Bought Items
1	{Bread, Cheese, Milk, Pasta}
2	{Bread, Cheese, Chips}
3	{Cheese, Pasta, Milk}
4	{Bread, Cheese, Milk}
5	{Bread, Pasta}

$$\text{conf}(\{\text{Bread}\} \Rightarrow \{\text{Cheese}, \text{Milk}\}) = \frac{\text{support}(\{\text{Bread}, \text{Cheese}, \text{Milk}\})}{\text{support}(\{\text{Bread}\})} = \frac{2}{4}$$

$$\text{conf}(\{\text{Cheese}, \text{Milk}\} \Rightarrow \{\text{Bread}\}) = \frac{\text{support}(\{\text{Bread}, \text{Cheese}, \text{Milk}\})}{\text{support}(\{\text{Cheese}, \text{Milk}\})} = \frac{2}{3}$$

Not symmetric (equality holds only in some rare cases)

Support and Confidence - Example

ID	Bought Items
1	{Bread, Cheese, Milk, Pasta}
2	{Bread, Cheese, Chips}
3	{Cheese, Pasta, Milk}
4	{Bread, Cheese, Milk}
5	{Bread, Pasta}

$$\text{conf}(\{\text{Bread}\} \Rightarrow \{\text{Cheese}, \text{Milk}\}) = \frac{\text{support}(\{\text{Bread},\text{Cheese},\text{Milk}\})}{\text{support}(\{\text{Bread}\})} = \frac{2}{4}$$

$$\text{conf}(\{\text{Cheese}, \text{Milk}\} \Rightarrow \{\text{Bread}\}) = \frac{\text{support}(\{\text{Bread},\text{Cheese},\text{Milk}\})}{\text{support}(\{\text{Cheese},\text{Milk}\})} = \frac{2}{3}$$

General rule:

$$\text{conf}(\{A, B\} \Rightarrow \{C\}) \geq \text{conf}(\{A\} \Rightarrow \{B, C\})$$

Probabilistic Interpretation

- Support: probability that an instance contains $\mathcal{A} \cup \mathcal{B}$

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \text{support}(\mathcal{A} \cup \mathcal{B}) \approx P(\mathcal{A} \cup \mathcal{B})$$

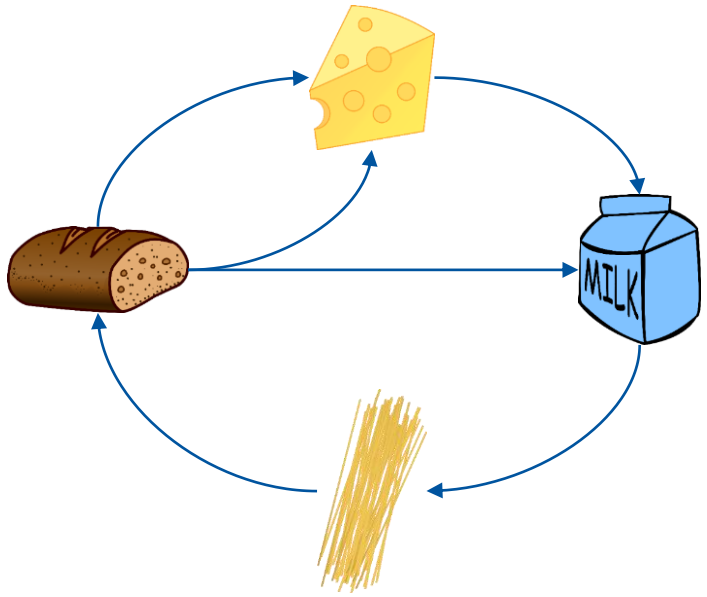
- Confidence: conditional probability that an instance contains items in $\mathcal{B}$, given that it contains items in $\mathcal{A}$

$$\text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\text{support}(\mathcal{A} \cup \mathcal{B})}{\text{support}(\mathcal{A})} \approx P(\mathcal{B} \mid \mathcal{A})$$

Take ‘probability’ with a grain of salt - we are only considering a sample.

Association Rules

1. Introduction
2. **Generating Association Rules**
3. Evaluation
4. Applications
5. Simpson’s Paradox



From Frequent Itemsets to Association Rules

Given: a dataset $\mathcal{X} \in \mathbb{M}(\mathbb{P}(\mathcal{I})), \text{min\_sup}, \text{min\_conf}$

How to generate all association rules that have high support and high confidence?

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \text{support}(\mathcal{A} \cup \mathcal{B}) \geq \text{min\_sup}$$

$$\text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\text{support}(\mathcal{A} \cup \mathcal{B})}{\text{support}(\mathcal{A})} \geq \text{min\_conf}$$

Frequent itemsets

Ensuring $\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) \geq \text{min\_sup}$

- ✓ Easy!
  - Use frequent itemsets as a basis
  - Consider frequent itemsets $\mathcal{C}$ such that $|\mathcal{C}| \geq 2$ and $\mathcal{C} \geq \text{min\_sup}$ (apply Apriori or FP-growth to generate such frequent itemsets)

Ensuring $\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) \geq \text{min\_sup}$

✓ Easy!

- Use frequent itemsets as a basis
- Consider frequent itemsets $\mathcal{C}$ such that $|\mathcal{C}| \geq 2$ and $\mathcal{C} \geq \text{min\_sup}$
  (apply Apriori or FP-growth to generate such frequent itemsets)
- Generate candidate rules $\mathcal{A} \Rightarrow \mathcal{B}$ by considering all splits of $\mathcal{C}$ into two non-empty disjoint subsets ($\mathcal{C} = \mathcal{A} \cup \mathcal{B}$)
- However: the number of such candidate rules is $2^{|\mathcal{C}|} - 2$!

Ensuring $\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) \geq \text{min\_sup}$

✓ Easy!

- Use frequent itemsets as a basis
- Consider frequent itemsets $\mathcal{C}$ such that $|\mathcal{C}| \geq 2$ and $\mathcal{C} \geq \text{min\_sup}$
  (apply Apriori or FP-growth to generate such frequent itemsets)
- Generate candidate rules $\mathcal{A} \Rightarrow \mathcal{B}$ by considering all splits of $\mathcal{C}$ into two non-empty disjoint subsets ($\mathcal{C} = \mathcal{A} \cup \mathcal{B}$)
- However: the number of such candidate rules is $2^{|\mathcal{C}|} - 2$!

$$\mathcal{C} = \{\{\text{Bread}\}, \{\text{Cheese}\}, \{\text{Milk}\}, \{\text{Pasta}\}\}$$

$\mathcal{A}$ $\mathcal{B}$

$\mathcal{A}$ $\mathcal{B}$

$\vdots$

$\mathcal{B}$ $\mathcal{A}$

$|\mathcal{C}| = 4 \implies 2^4 - 2 = 14$ candidate rules

… and the number of candidate frequent itemsets was already exponential!

Ensuring $\mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B}) \geq \mathrm{min\_conf}$

No additional pass over the data needed

- Itemsets $\mathcal{A} \cup \mathcal{B}$ and $\mathcal{A}$ are frequent
  → their supports have already been computed when using Apriori or FP-growth
- Therefore, we can simply test every candidate rule and only return the ones that satisfy the criterion:

$$\mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\mathrm{support}(\mathcal{A} \cup \mathcal{B})}{\mathrm{support}(\mathcal{A})} \geq \mathrm{min\_conf}$$

Ensuring $\mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B}) \geq \mathrm{min\_conf}$

- Itemsets $\mathcal{A} \cup \mathcal{B}$ and $\mathcal{A}$ are frequent
  → their supports have already been computed when using Apriori or FP-growth
- Therefore, we can simply test every candidate rule and only return the ones that satisfy the criterion:

$$\mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\mathrm{support}(\mathcal{A} \cup \mathcal{B})}{\mathrm{support}(\mathcal{A})} \geq \mathrm{min\_conf}$$

But…
- There could be way too many association rules.
- Most are not interesting!

Confidence-Based Pruning

- Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$, and itemset $\mathcal{C}$ such that $\mathcal{C} \cap \mathcal{A} = \emptyset$
- It holds that $\text{conf}(\mathcal{A} \Rightarrow \mathcal{B} \cup \mathcal{C}) = \frac{\text{support}(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C})}{\text{support}(\mathcal{A})} \leq \frac{\text{support}(\mathcal{A} \cup \mathcal{B})}{\text{support}(\mathcal{A})} = \text{conf}(\mathcal{A} \Rightarrow \mathcal{B})$

recall that the support of a superset is lower or equal

Confidence-Based Pruning

- Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$, and itemset $\mathcal{C}$ such that $\mathcal{C} \cap \mathcal{A} = \emptyset$
- It holds that $\mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B} \cup \mathcal{C}) = \frac{\mathrm{support}(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C})}{\mathrm{support}(\mathcal{A})} \leq \frac{\mathrm{support}(\mathcal{A} \cup \mathcal{B})}{\mathrm{support}(\mathcal{A})} = \mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B})$

- Hence, if $\mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B}) \leq \mathrm{min\_conf}$ then $\mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B} \cup \mathcal{C}) \leq \mathrm{min\_conf}$
- Adding $\mathcal{C}$ to the right part makes the rule stronger
- We can focus on the strongest rules meeting the confidence threshold

Confidence-Based Pruning

- Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$, and itemset $\mathcal{C}$ such that $\mathcal{C} \cap \mathcal{A} = \emptyset$
- It holds that $\text{conf}(\mathcal{A} \Rightarrow \mathcal{B} \cup \mathcal{C}) = \frac{\text{support}(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C})}{\text{support}(\mathcal{A})} \leq \frac{\text{support}(\mathcal{A} \cup \mathcal{B})}{\text{support}(\mathcal{A})} = \text{conf}(\mathcal{A} \Rightarrow \mathcal{B})$

- Hence, if $\text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) \leq \text{min\_conf}$ then $\text{conf}(\mathcal{A} \Rightarrow \mathcal{B} \cup \mathcal{C}) \leq \text{min\_conf}$
- Adding $\mathcal{C}$ to the right part makes the rule stronger
- We can focus on the strongest rules meeting the confidence threshold

- No clear relation between $\text{conf}(\mathcal{A} \cup \mathcal{C} \Rightarrow \mathcal{B})$ and $\text{conf}(\mathcal{A} \Rightarrow \mathcal{B})$

Not even for $\mathcal{C} \cap (\mathcal{A} \cup \mathcal{B}) = \emptyset$

- Additions to the left part of the rule may lead to an increase or decrease
  - $\{\text{Cheese}\} \Rightarrow \{\text{Wine}\}$ may have a confidence of 0.2
  - $\{\text{Cheese}, \text{Babyfood}\} \Rightarrow \{\text{Wine}\}$ may have a confidence of 0.1
  - $\{\text{Cheese}, \text{Chips}\} \Rightarrow \{\text{Wine}\}$ may have a confidence of 0.3

Redundant Rules

- Consider two different association rules $\mathcal{A} \Rightarrow \mathcal{B}$ and $\mathcal{A}' \Rightarrow \mathcal{B}'$ with identical support and confidence, i.e.:
  - $\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \text{support}(\mathcal{A}' \Rightarrow \mathcal{B}')$
  - $\text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \text{conf}(\mathcal{A}' \Rightarrow \mathcal{B}')$

- $\mathcal{A}' \Rightarrow \mathcal{B}'$ is redundant if $\mathcal{A}' \subseteq \mathcal{A}$ and $\mathcal{B}' \subseteq \mathcal{B}$

- Using only closed frequent itemsets will avoid generating redundant rules

  (Recall: An itemset is closed if there is no proper superset that has the same support)

Avoiding Generation of Redundant Rules

1. Assume $\mathcal{A}' \Rightarrow \mathcal{B}'$ is redundant, i.e., there is another rule $\mathcal{A} \Rightarrow \mathcal{B}$ such that
   - $\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \text{support}(\mathcal{A}' \Rightarrow \mathcal{B}')$
   - $\text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \text{conf}(\mathcal{A}' \Rightarrow \mathcal{B}')$
   - $\mathcal{A}' \subseteq \mathcal{A}$
   - $\mathcal{B}' \subseteq \mathcal{B}$
   - It holds that $\mathcal{A}' \cup \mathcal{B}' \subset \mathcal{A} \cup \mathcal{B}$ (because the rules are different)

Avoiding Generation of Redundant Rules

1. Assume $\mathcal{A}' \Rightarrow \mathcal{B}'$ is redundant, i.e., there is another rule $\mathcal{A} \Rightarrow \mathcal{B}$ such that
   - $\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \text{support}(\mathcal{A}' \Rightarrow \mathcal{B}')$
   - $\text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \text{conf}(\mathcal{A}' \Rightarrow \mathcal{B}')$
   - $\mathcal{A}' \subseteq \mathcal{A}$
   - $\mathcal{B}' \subseteq \mathcal{B}$
   - It holds that $\mathcal{A}' \cup \mathcal{B}' \subset \mathcal{A} \cup \mathcal{B}$ (because the rules are different)
2. Also, assume $\mathcal{A} \cup \mathcal{B}$ and $\mathcal{A}' \cup \mathcal{B}'$ are closed, i.e., there are no proper supersets with the same support
   - Hence, $\text{support}(\mathcal{A}' \Rightarrow \mathcal{B}') > \text{support}(\mathcal{A} \Rightarrow \mathcal{B})$ (cannot be equal, $\mathcal{A} \cup \mathcal{B}$ is closed)

Therefore, we find a contradiction.

Closed itemsets cannot produce redundant rules.

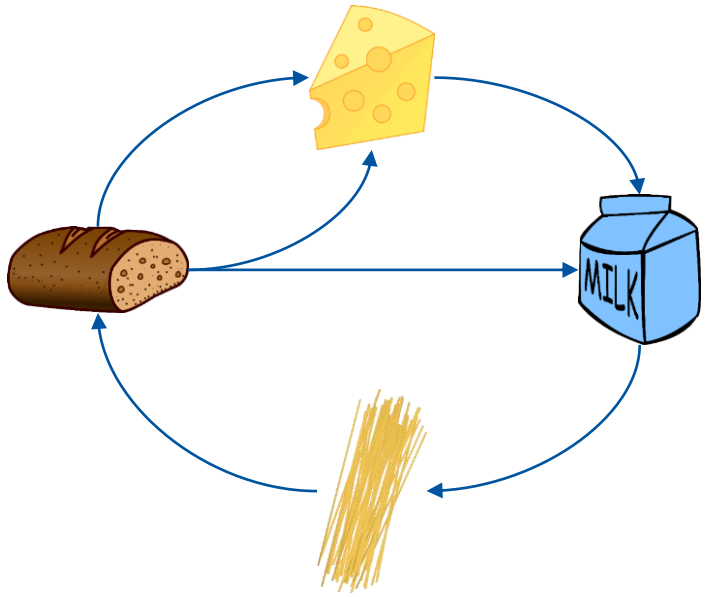
Summary

How to generate association rules that are interesting?

- We can generate candidate rules with high support based on frequent itemsets
- We can filter those candidates with high confidence without going back to the data
- We can prune the rules based on confidence: $\text{min\_conf} \leq \text{conf}(\mathcal{A} \Rightarrow \mathcal{B} \cup \mathcal{C}) \leq \cancel{\text{conf}(\mathcal{A} \Rightarrow \mathcal{B})}$
- We can focus on closed frequent itemsets to avoid redundant rules
- Not enough: we need additional evaluation concepts such as “surprisingness” (lift)

Association Rules

1. Introduction
2. Generating Association Rules
3. **Evaluation**
4. Applications
5. Simpson's Paradox



Association rules

- {Cheese, Chips} ⇒ {Wine, Beer}
- {One(Metallica),Trasher(Evile)} ⇒ {Augen-Auf(Oomph), The Trooper(Iron Maiden)}
- {Temp>20,Play=Yes} ⇒ {Wind=No}
- {Night_flight=No, Traffic=Yes} ⇒ {Flight_delayed=Yes}
- {Gender=Male, Sport=Football} ⇒ {Favorite_food=Currywurst, Age>40}
- …

How to evaluate the quality of a rule?

Confusion matrix for association rules

Consider a set of transactions and an association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	$\#\mathcal{A}\mathcal{B}$	$\#\mathcal{A}\overline{\mathcal{B}}$	$\#\mathcal{A}$
$\mathcal{A}$ is not included	$\#\overline{\mathcal{A}}\mathcal{B}$	$\#\overline{\mathcal{A}\mathcal{B}}$	$\#\overline{\mathcal{A}}$
	$\#\mathcal{B}$	$\#\overline{\mathcal{B}}$	$\#\mathrm{ALL}$

Number of transactions with $\mathcal{A}$ included but $\mathcal{B}$ not included

Bottom row: Sum of the two cells above

Rightmost column: sum of the two cells to the left

Confusion matrix for association rules

Consider a set of transactions and an association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	$\#\mathcal{AB}$	$\#\mathcal{A}\overline{\mathcal{B}}$	$\#\mathcal{A}$
$\mathcal{A}$ is not included	$\#\overline{\mathcal{A}}\mathcal{B}$	$\#\overline{\mathcal{AB}}$	$\#\overline{\mathcal{A}}$
	$\#\mathcal{B}$	$\#\overline{\mathcal{B}}$	$\#\mathrm{ALL}$

$$\mathrm{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\#\mathcal{AB}}{\#ALL} \qquad \mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\#\mathcal{AB}}{\#\mathcal{A}}$$

Confusion matrix for association rules

Consider a set of transactions and an association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	$\#\mathcal{AB}$	$\#\mathcal{A}\overline{\mathcal{B}}$	$\#\mathcal{A}$
$\mathcal{A}$ is not included	$\#\overline{\mathcal{A}}\mathcal{B}$	$\#\overline{\mathcal{AB}}$	$\#\overline{\mathcal{A}}$
	$\#\mathcal{B}$	$\#\overline{\mathcal{B}}$	$\#\mathrm{ALL}$

The lower the better

The higher the better

Not captured in any of the metrics

$$\mathrm{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\#\mathcal{AB}}{\#ALL} \qquad \mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\#\mathcal{AB}}{\#\mathcal{A}}$$

High Support and High Confidence

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	100	0	100
$\mathcal{A}$ is not included	0	0	0
	100	0	100

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{100}{100} \qquad \text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{100}{100}$$

Low Support and High Confidence

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	10	0	10
$\mathcal{A}$ is not included	40	50	90
	50	50	100

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{10}{100} \qquad \text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{10}{10}$$

Low Support and Low Confidence

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	10	40	50
$\mathcal{A}$ is not included	25	25	50
	35	65	100

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{10}{100} \qquad \text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{10}{50}$$

Support and Confidence Don't Tell The Full Story

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	80	10	90
$\mathcal{A}$ is not included	**0**	**10**	10
	80	20	100

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{80}{100} \qquad \text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{80}{90}$$

Seems to be a good rule because if $\mathcal{A}$ is not included, $\mathcal{B}$ is also never included

Support and Confidence Don't Tell The Full Story

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	80	10	90
$\mathcal{A}$ is not included	**10**	**0**	10
	90	10	100

The distribution of counts in the second row does not influence support and confidence

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{80}{100} \qquad \text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{80}{90}$$

Same support and confidence, but seems to be a poor rule because if $\mathcal{A}$ is not included, $\mathcal{B}$ is always included

We need Lift: How surprising?

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	$\#\mathcal{A}\mathcal{B}$	$\#\mathcal{A}\overline{\mathcal{B}}$	$\#\mathcal{A}$
$\mathcal{A}$ is not included	$\#\overline{\mathcal{A}}\mathcal{B}$	$\#\overline{\mathcal{A}\mathcal{B}}$	$\#\overline{\mathcal{A}}$
	$\#\mathcal{B}$	$\#\overline{\mathcal{B}}$	$\#\mathrm{ALL}$

$$\mathrm{lift}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\mathrm{support}(\mathcal{A} \cup \mathcal{B})}{\mathrm{support}(\mathcal{A}) \cdot \mathrm{support}(\mathcal{B})} = \frac{P(\mathcal{A} \cup \mathcal{B})}{P(\mathcal{A}) \cdot P(\mathcal{B})} = \frac{\frac{\#\mathcal{A}\mathcal{B}}{\#\mathrm{ALL}}}{\frac{\#\mathcal{A}}{\#\mathrm{ALL}} \cdot \frac{\#\mathcal{B}}{\#\mathrm{ALL}}} = \frac{\#\mathcal{A}\mathcal{B} \cdot \#\mathrm{ALL}}{\#\mathcal{A} \cdot \#\mathcal{B}}$$

We need Lift: How surprising?

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$$\text{lift}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\text{support}(\mathcal{A} \cup \mathcal{B})}{\text{support}(\mathcal{A}) \cdot \text{support}(\mathcal{B})} = \frac{P(\mathcal{A} \cup \mathcal{B})}{P(\mathcal{A}) \cdot P(\mathcal{B})} = \frac{\frac{\#\mathcal{A}\mathcal{B}}{\#\text{ALL}}}{\frac{\#\mathcal{A}}{\#\text{ALL}} \cdot \frac{\#\mathcal{B}}{\#\text{ALL}}} = \frac{\#\mathcal{A}\mathcal{B} \cdot \#\text{ALL}}{\#\mathcal{A} \cdot \#\mathcal{B}}$$

If $\text{lift}(\mathcal{A} \Rightarrow \mathcal{B}) \approx 1$ then $\mathcal{A}$ and $\mathcal{B}$ are **independent** $P(\mathcal{A} \cup \mathcal{B}) \approx P(\mathcal{A}) \cdot P(\mathcal{B})$

If $\text{lift}(\mathcal{A} \Rightarrow \mathcal{B}) \gg 1$ then $\mathcal{A}$ and $\mathcal{B}$ are **positively correlated** $P(\mathcal{A} \cup \mathcal{B}) \gg P(\mathcal{A}) \cdot P(\mathcal{B})$

$\mathcal{A}$ and $\mathcal{B}$ are likely to be included together

If $\text{lift}(\mathcal{A} \Rightarrow \mathcal{B}) \ll 1$ then $\mathcal{A}$ and $\mathcal{B}$ are **negatively correlated** $P(\mathcal{A} \cup \mathcal{B}) \ll P(\mathcal{A}) \cdot P(\mathcal{B})$

$\mathcal{A}$ and $\mathcal{B}$ are unlikely to be included together (consider $\mathcal{A} \Rightarrow \neg\mathcal{B}$)

We need Lift

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	$\#\mathcal{A}\mathcal{B}$	$\#\mathcal{A}\overline{\mathcal{B}}$	$\#\mathcal{A}$
$\mathcal{A}$ is not included	$\#\overline{\mathcal{A}}\mathcal{B}$	$\#\overline{\mathcal{A}\mathcal{B}}$	$\#\overline{\mathcal{A}}$
	$\#\mathcal{B}$	$\#\overline{\mathcal{B}}$	$\#\mathrm{ALL}$

$$\mathrm{lift}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\mathrm{support}(\mathcal{A} \cup \mathcal{B})}{\mathrm{support}(\mathcal{A}) \cdot \mathrm{support}(\mathcal{B})} = \frac{P(\mathcal{A} \cup \mathcal{B})}{P(\mathcal{A}) \cdot P(\mathcal{B})} = \frac{\frac{\#\mathcal{A}\mathcal{B}}{\#\mathrm{ALL}}}{\frac{\#\mathcal{A}}{\#\mathrm{ALL}} \cdot \frac{\#\mathcal{B}}{\#\mathrm{ALL}}}$$

If $\mathrm{lift}(\mathcal{A} \Rightarrow \mathcal{B}) \approx 1$ then $\mathcal{A}$ and $\mathcal{B}$ are **independent** $P(\mathcal{A} \cup \mathcal{B}) \approx P(\mathcal{A}) \cdot P(\mathcal{B})$

If $\mathrm{lift}(\mathcal{A} \Rightarrow \mathcal{B}) \ll 1$ then $\mathcal{A}$ and $\mathcal{B}$ are **negatively correlated** $P(\mathcal{A} \cup \mathcal{B}) \ll P(\mathcal{A}) \cdot P(\mathcal{B})$

If $\mathrm{lift}(\mathcal{A} \Rightarrow \mathcal{B}) \gg 1$ then $\mathcal{A}$ and $\mathcal{B}$ are **positively correlated** $P(\mathcal{A} \cup \mathcal{B}) \gg P(\mathcal{A}) \cdot P(\mathcal{B})$

Is the Rule Surprising?

Different frequencies in the rows, but distribution is the same

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	9	1	10
$\mathcal{A}$ is not included	81	9	90
	90	10	100

$$\text{lift}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\text{support}(\mathcal{A} \cup \mathcal{B})}{\text{support}(\mathcal{A}) \cdot \text{support}(\mathcal{B})} = \frac{P(\mathcal{A} \cup \mathcal{B})}{P(\mathcal{A}) \cdot P(\mathcal{B})} = \frac{\frac{\#\mathcal{A}\mathcal{B}}{\#\text{ALL}}}{\frac{\#\mathcal{A}}{\#\text{ALL}} \cdot \frac{\#\mathcal{B}}{\#\text{ALL}}}$$

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{9}{100}$$

$$\text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{9}{10}$$

$$\text{lift}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\frac{9}{100}}{\frac{10}{100} \cdot \frac{90}{100}} = 1$$

No suprise!

Is the Rule Surprising?

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	9	1	10
$\mathcal{A}$ is not included	**0**	**90**	90
	9	91	100

$$\text{lift}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\text{support}(\mathcal{A} \cup \mathcal{B})}{\text{support}(\mathcal{A}) \cdot \text{support}(\mathcal{B})} = \frac{P(\mathcal{A} \cup \mathcal{B})}{P(\mathcal{A}) \cdot P(\mathcal{B})} = \frac{\frac{\#\mathcal{A}\mathcal{B}}{\#\text{ALL}}}{\frac{\#\mathcal{A}}{\#\text{ALL}} \cdot \frac{\#\mathcal{B}}{\#\text{ALL}}}$$

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{9}{100}$$

$$\text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{9}{10}$$

$$\text{lift}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\frac{9}{100}}{\frac{10}{100} \cdot \frac{9}{100}} = 10$$

Surprise!

Is the Rule Surprising?

Consider association rule $\mathcal{A} \Rightarrow \mathcal{B}$

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	9	1	10
$\mathcal{A}$ is not included	**90**	**0**	90
	99	1	100

$$\text{lift}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\text{support}(\mathcal{A} \cup \mathcal{B})}{\text{support}(\mathcal{A}) \cdot \text{support}(\mathcal{B})} = \frac{P(\mathcal{A} \cup \mathcal{B})}{P(\mathcal{A}) \cdot P(\mathcal{B})} = \frac{\frac{\#\mathcal{A}\mathcal{B}}{\#\text{ALL}}}{\frac{\#\mathcal{A}}{\#\text{ALL}} \cdot \frac{\#\mathcal{B}}{\#\text{ALL}}}$$

$$\text{support}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{9}{100}$$

$$\text{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{9}{10}$$

$$\text{lift}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\frac{9}{100}}{\frac{10}{100} \cdot \frac{99}{100}} = \frac{10}{11}$$

a little bit …

Selecting Association rules

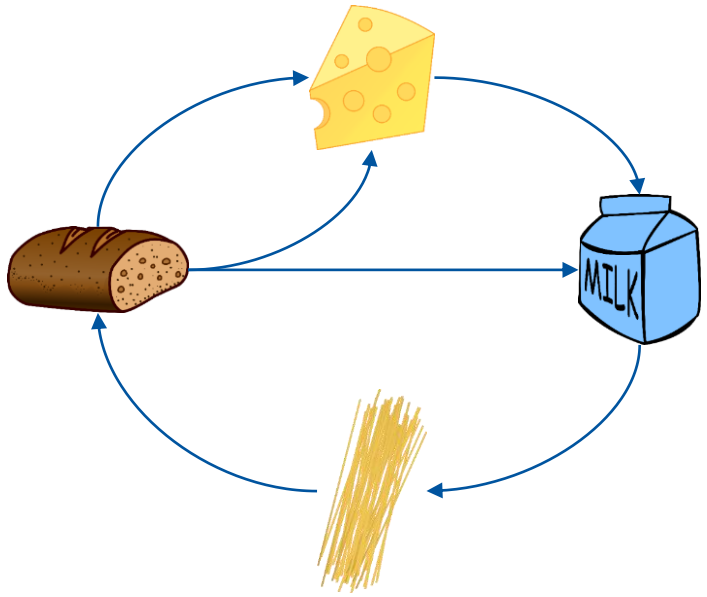
1. Set thresholds for minimal support and confidence
2. Evaluate lift and possibly other metrics for the rules remaining
3. Sort and prune based on any of the quality criteria (support, confidence, lift, etc.)

It is hard to predict the number of rules beforehand

There are many other measures of quality (conviction, leverage, collective strength, etc.)

Association Rules

1. Introduction
2. Generating Association Rules
3. Evaluation
4. **Applications**
5. Simpson's Paradox



Spotify

{Flowers(Miley Cyrus), Unholy(Sam Smith)} ⇒ {Levitating(Dua Lipa)}

{One(Metallica),Trasher(Evile)} ⇒ {Augen-Auf(Oomph),The Trooper(Iron Maiden)}

{Birds(Anouk),Irgendwo(Nena)} ⇒ {Leiser(Lea),Klavier(Lea)}

- 456 million active listeners
- 195 million premium subscribers
- Over 80 million songs

(As of January 2023)

Amazon

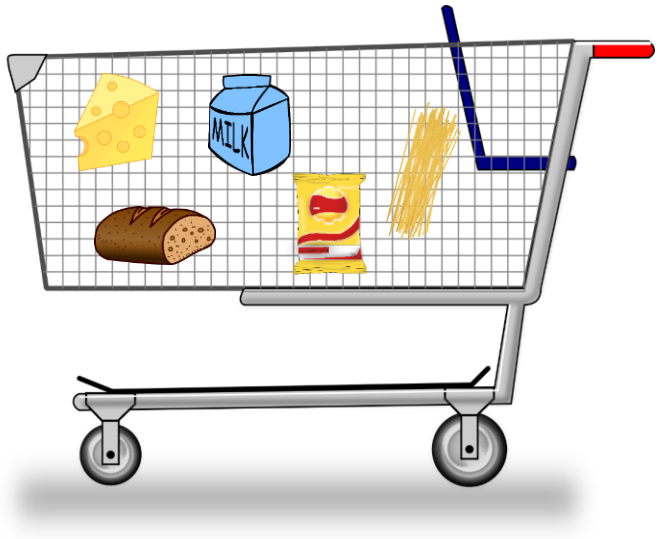
{Echo-Show-8,Fire-TV-Cube} ⇒ {Kindle-Paperwhite}

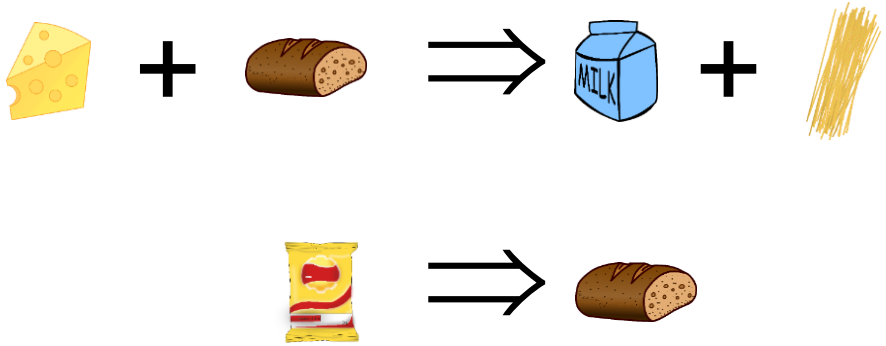
{Fire-TV-Stick-8} ⇒ {Fire-HD-8,Blink-Mini}

- 300 million active users
- Over 2 million third-party seller businesses
- Around 350 million items on the marketplace

(As of January 2023)

Supermarkets





support = 0.01
confidence = 0.85
lift = 1.67

support = 0.001
confidence = 0.15
lift = 1.2

Using Features Values As Items And Instances As Itemsets

Rain	Wind	Temp	Play
Yes	Yes	15	No
No	No	34	Yes
Yes	No	23	Yes
Yes	Yes	20	Yes
No	Yes	28	No
…	…	…	…

- Examples consider items as products, services, etc.
- Items can also be normal features values and transactions normal instances
- This leads to itemsets of the form $\{f_1=v_1, f_2=v_2, \ldots f_n=v_n\}$ for each instance

Using Features Values As Items And Instances As Itemsets

Rain	Wind	Temp	Play
Yes	Yes	15	No
No	No	34	Yes
Yes	No	23	Yes
Yes	Yes	20	Yes
No	Yes	28	No
…	…	…	…

[{Rain=Yes, Wind=Yes, Temp=15, Play=No},
{Rain=No, Wind=No, Temp=34, Play=Yes},
{Rain=Yes, Wind=No, Temp=23, Play=Yes},
{Rain=Yes, Wind=Yes, Temp=20, Play=Yes},
{Rain=No, Wind=Yes, Temp=28, Play=No},
…]

- Examples consider items as products, services, etc.
- Items can also be normal features values and transactions normal instances
- This leads to itemsets of the form $\{f_1=v_1, f_2=v_2, \ldots f_n=v_n\}$ for each instance

Using Features Values As Items And Instances As Itemsets

Rain	Wind	Temp	Play
Yes	Yes	15	No
No	No	34	Yes
Yes	No	23	Yes
Yes	Yes	20	Yes
No	Yes	28	No
…	…	…	…

[{Rain=Yes, Wind=Yes, 10≤Temp<20, Play=No},
{Rain=No, Wind=No, 30≤Temp<40, Play=Yes},
{Rain=Yes, Wind=No, 20≤Temp<30, Play=Yes},
{Rain=Yes, Wind=Yes, 20≤Temp<30, Play=Yes},
{Rain=No, Wind=Yes, 20≤Temp<30, Play=No},
…]

- Items can also be ranges for continuous feature values
  - Temp≥25
  - Temp<25
  - 20≤Temp<30
  - Etc.
- Any dataset having instances and features can be converted into a multiset of transactions $\mathcal{X} \in \mathbb{M}(\mathbb{P}(\mathcal{I}))$

Using Features Values As Items And Instances As Itemsets

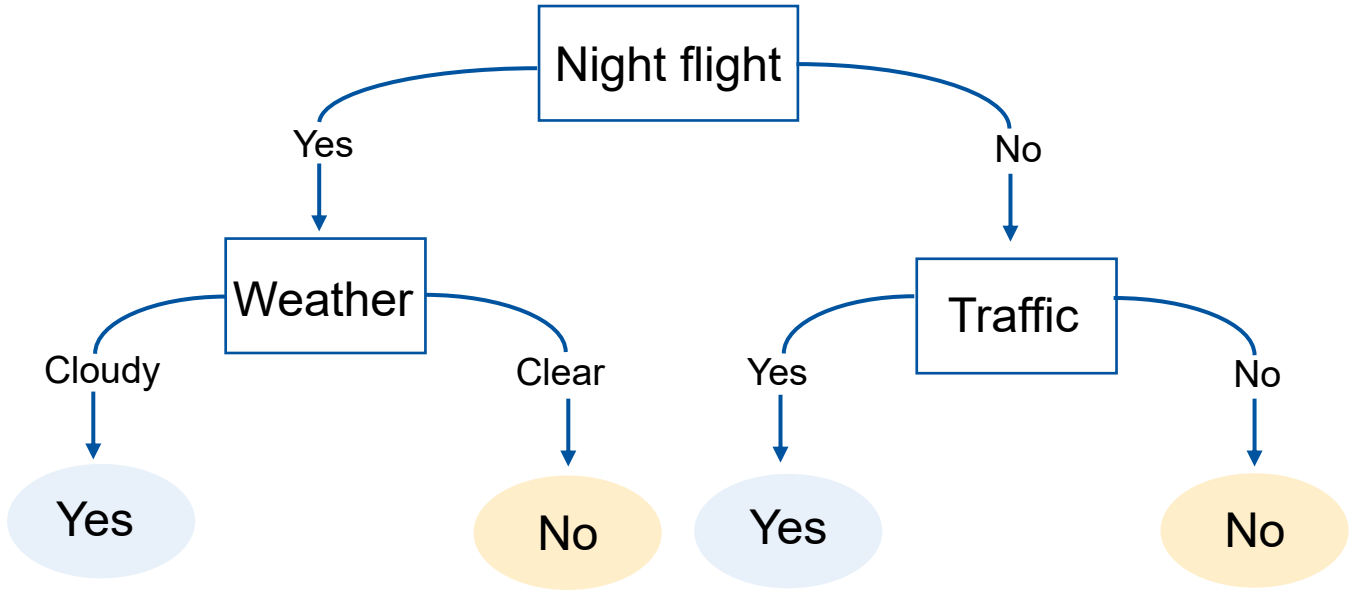
Rain	Wind	Temp	Play
Yes	Yes	15	No
No	No	34	Yes
Yes	No	23	Yes
Yes	Yes	20	Yes
No	Yes	28	No
…	…	…	…

{Rain=Yes, Wind=Yes} $\Rightarrow$ {Play=No}
{Temp>30} $\Rightarrow$ {Rain=No, Wind=No}
{Temp>20,Play=Yes} $\Rightarrow$ {Wind=No}

- Any dataset having instances and features can be converted into a multiset of transactions $\mathcal{X} \in \mathbb{M}(\mathbb{P}(\mathcal{I}))$
- Hence, we can also have association rules of the form
  $\mathcal{A} \Rightarrow \mathcal{B}$ with $\mathcal{A} \subseteq \mathcal{I}, \mathcal{B} \subseteq \mathcal{I}$ and $\mathcal{A} \cap \mathcal{B} = \emptyset$

Link To Classification and Decision Trees

Weather	Traffic	Night flight	Flight delayed
Cloudy	No	Yes	**Yes**
Cloudy	Yes	No	**Yes**
Cloudy	Yes	No	**Yes**
Clear	Yes	Yes	**No**
Clear	No	No	**No**
Clear	No	No	**No**



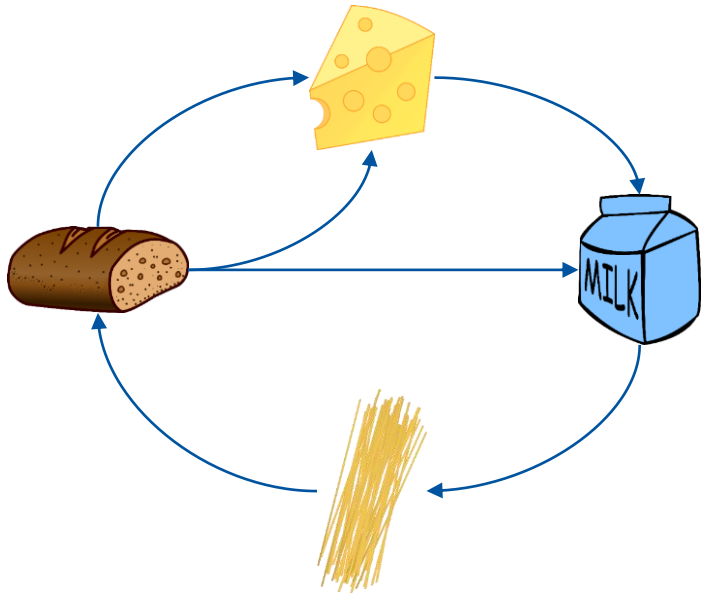
{Night_flight=Yes, Weather=Cloudy} ⇒ {Flight_delayed=Yes}
{Night_flight=Yes, Weather=Clear} ⇒ {Flight_delayed=No}
{Night_flight=No, Traffic=Yes} ⇒ {Flight_delayed=Yes}
{Night_flight=No, Traffic=No} ⇒ {Flight_delayed=No}

Summary

- Association rules can be learned for “normal itemsets” and itemsets based on feature values
- Classification rules can be expressed as association rules
- The challenge remains that there are exponentially many candidate rules
- Confidence and support are only part of the story
  - What if many rules meet the two thresholds?
  - How to select the most interesting ones?
  - Many more quality metrics exist, such as lift

Association Rules

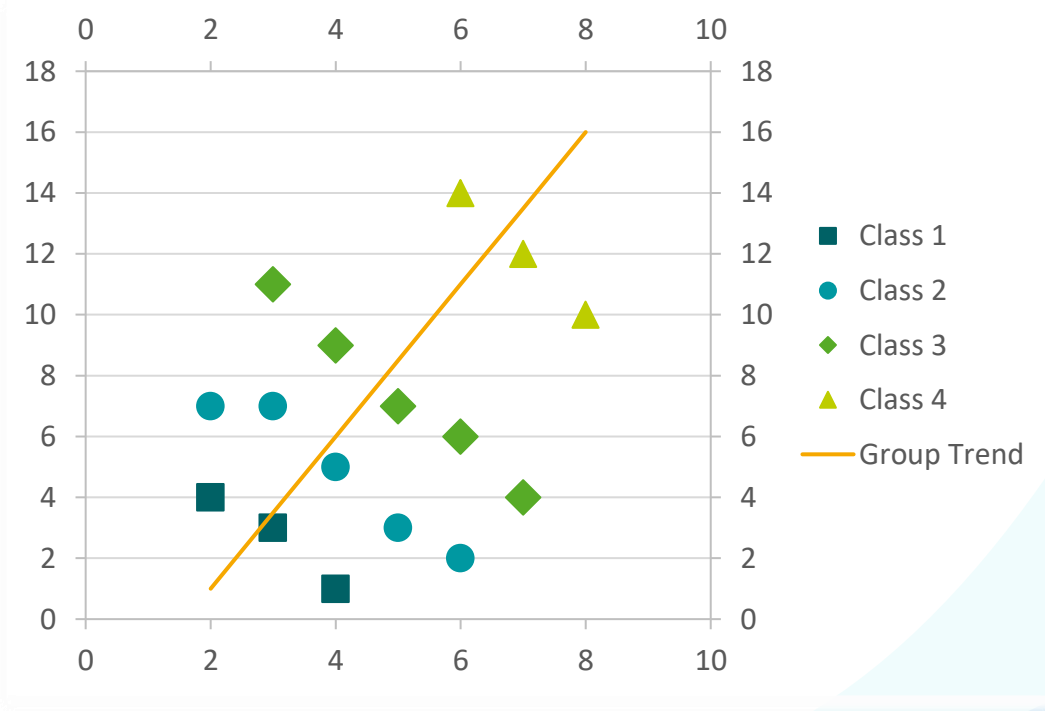
1. Introduction
2. Generating Association Rules
3. Evaluation
4. Applications
5. **Simpson's Paradox**



Simpson’s Paradox

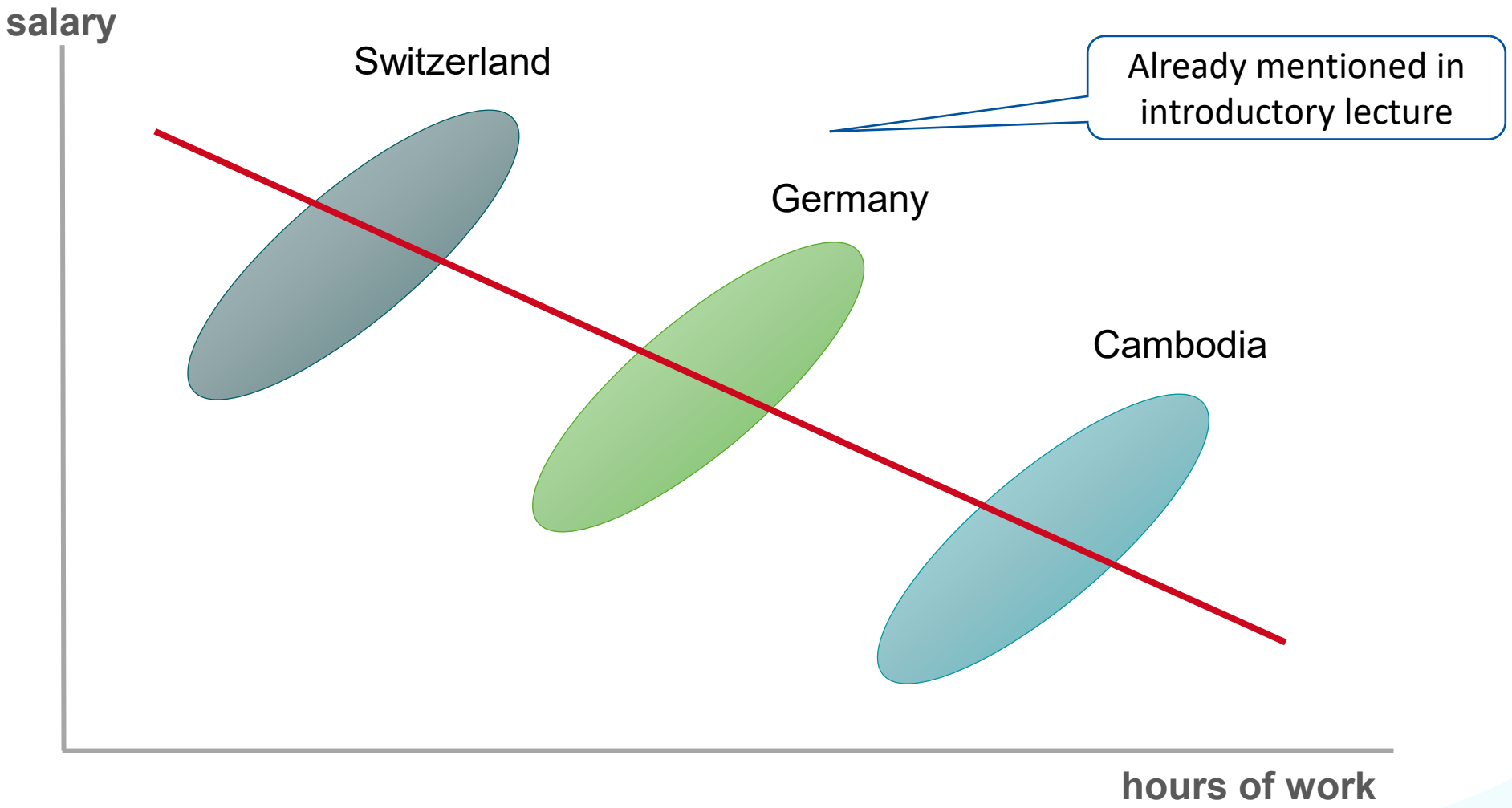
A trend appears in several different groups of data but disappears or reverses when these groups are combined.

Already mentioned in introductory lecture



- Edward Simpson in 1951 (earlier variants by Udny Yule and Karl Pearson)
- Nice example of ‘How to lie with statistics?’
- The paradox is often encountered in social-science and medical-science

Simpson’s Paradox When Using Regression



Simpson's Paradox in Association Rules

Consider the association rule $\mathcal{A} \Rightarrow \mathcal{B}$ and any feature which splits the instances (location, age …)

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	$a + p$	$(b - a) + (q - p)$	$b + q$
$\mathcal{A}$ is not included	$c + r$	$(d - c) + (s - r)$	$d + s$
	$a + c + p + r$	$(b + d + q + s) - (a + c + p + r)$	$b + d + q + s$

Two classes – blue and orange
(e.g., old and young)

Simpson's Paradox in Association Rules

Consider the association rule $\mathcal{A} \Rightarrow \mathcal{B}$ and any feature which splits the instances (location, age …)

$\mathcal{A} \Rightarrow \mathcal{B}$	$\mathcal{B}$ is included	$\mathcal{B}$ is not included	
$\mathcal{A}$ is included	$a + p$	$(b - a) + (q - p)$	$b + q$
$\mathcal{A}$ is not included	$c + r$	$(d - c) + (s - r)$	$d + s$
	$a + c + p + r$	$(b + d + q + s) - (a + c + p + r)$	$b + d + q + s$

$$\mathrm{conf}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{a+p}{b+q}$$

$$\mathrm{lift}(\mathcal{A} \Rightarrow \mathcal{B}) = \frac{\frac{a+p}{b+d+q+s}}{\frac{b+q}{b+d+q+s} \cdot \frac{a+c+p+r}{b+d+q+s}} = \frac{(a+p)\cdot(b+d+q+s)}{(b+q)\cdot(a+c+p+r)}$$

Simpson's Paradox - Example

Two classes: old and young

smoke $\Rightarrow$ cancer	has cancer	doesn't have cancer	
smokes	$1 + 66$	$2 + 34$	$3 + 100$
doesn't smoke	$34 + 2$	$66 + 1$	$100 + 3$
	$35 + 68$	$68 + 35$	$103 + 103$

humans $\quad \text{conf}(\text{smoke} \Rightarrow \text{cancer}) = \frac{67}{103} = 0.65 > \text{conf}(\text{not smoke} \Rightarrow \text{cancer}) = \frac{36}{103} = 0.35$

old $\quad \text{conf}(\text{smoke} \Rightarrow \text{cancer}) = \frac{1}{3} = 0.333 < \text{conf}(\text{not smoke} \Rightarrow \text{cancer}) = \frac{34}{100} = 0.34$

young $\quad \text{conf}(\text{smoke} \Rightarrow \text{cancer}) = \frac{66}{100} = 0.66 < \text{conf}(\text{not smoke} \Rightarrow \text{cancer}) = \frac{2}{3} = 0.666$

Simpson's Paradox - Example

Two classes: old and young

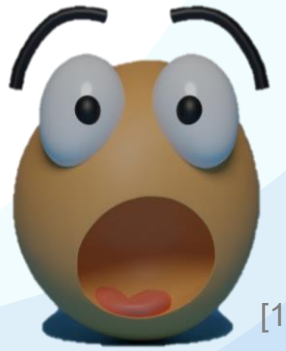
smoke $\Rightarrow$ cancer	has cancer	doesn't have cancer	
smokes	$1 + 66$	$2 + 34$	$3 + 100$
doesn't smoke	$34 + 2$	$66 + 1$	$100 + 3$
	$35 + 68$	$68 + 35$	$103 + 103$

humans $\quad \text{conf}(\text{smoke} \Rightarrow \text{cancer}) = \frac{67}{103} = 0.65 > \text{conf}(\text{not smoke} \Rightarrow \text{cancer}) = \frac{36}{103} = 0.35$

old $\quad \text{conf}(\text{smoke} \Rightarrow \text{cancer}) = \frac{1}{3} = 0.333 < \text{conf}(\text{not smoke} \Rightarrow \text{cancer}) = \frac{34}{100} = 0.34$

young $\quad \text{conf}(\text{smoke} \Rightarrow \text{cancer}) = \frac{66}{100} = 0.66 < \text{conf}(\text{not smoke} \Rightarrow \text{cancer}) = \frac{2}{3} = 0.666$

Smoking is healthy for old and young people, but not for all humans!



[1]

Simpson's Paradox - Example

Two classes: old and young

smoke $\Rightarrow$ cancer	has cancer	doesn't have cancer	
smokes	$1 + 66$	$2 + 34$	$3 + 100$
doesn't smoke	$34 + 2$	$66 + 1$	$100 + 3$
	$35 + 68$	$68 + 35$	$103 + 103$

- The presence of smoking has a strong positive effect on the occurrence of cancer in the overall set (supports the rule)
- However, the effect cannot be seen in the subsets!

Simpson's Paradox - Example

Two classes: old and young

smoke $\Rightarrow$ cancer	has cancer	doesn't have cancer	
smokes	$1 + 66$	$2 + 34$	$3 + 100$
doesn't smoke	$34 + 2$	$66 + 1$	$100 + 3$
	$35 + 68$	$68 + 35$	$103 + 103$

humans $\quad \text{lift}(\text{smoke} \Rightarrow \text{cancer}) = \frac{\frac{67}{206}}{\frac{103}{206} \cdot \frac{103}{206}} = \frac{67 \cdot 206}{103 \cdot 103} = 1.301$

old $\quad \text{lift}(\text{smoke} \Rightarrow \text{cancer}) = \frac{\frac{1}{103}}{\frac{3}{103} \cdot \frac{35}{103}} = \frac{1 \cdot 103}{3 \cdot 35} = 0.9809$

young $\quad \text{lift}(\text{smoke} \Rightarrow \text{cancer}) = \frac{\frac{66}{103}}{\frac{100}{103} \cdot \frac{68}{103}} = \frac{66 \cdot 103}{100 \cdot 68} = 0.9997$

Simpson's Paradox - Example

Two classes: old and young

smoke $\Rightarrow$ cancer	has cancer	doesn't have cancer	
smokes	$1 + 66$	$2 + 34$	$3 + 100$
doesn't smoke	$34 + 2$	$66 + 1$	$100 + 3$
	$35 + 68$	$68 + 35$	$103 + 103$

humans $\quad \text{lift}(\text{smoke} \Rightarrow \text{cancer}) = \frac{\frac{67}{206}}{\frac{103}{206} \cdot \frac{103}{206}} = \frac{67 \cdot 206}{103 \cdot 103} = 1.301$

Positively correlated

old $\quad \text{lift}(\text{smoke} \Rightarrow \text{cancer}) = \frac{\frac{1}{103}}{\frac{3}{103} \cdot \frac{35}{103}} = \frac{1 \cdot 103}{3 \cdot 35} = 0.9809$

Negatively correlated

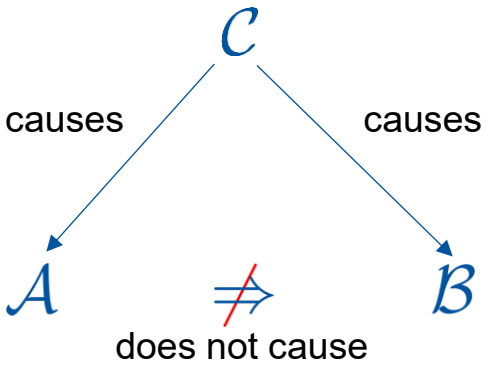
young $\quad \text{lift}(\text{smoke} \Rightarrow \text{cancer}) = \frac{\frac{66}{103}}{\frac{100}{103} \cdot \frac{68}{103}} = \frac{66 \cdot 103}{100 \cdot 68} = 0.9997$

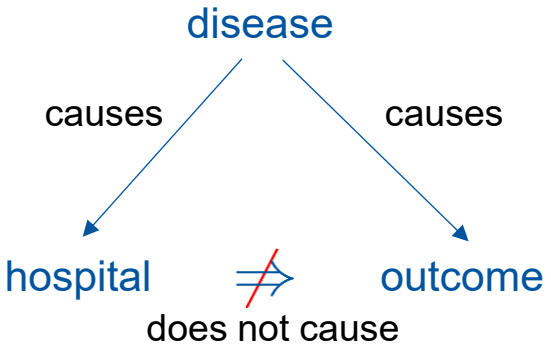
Simpson’s Paradox – Other Examples

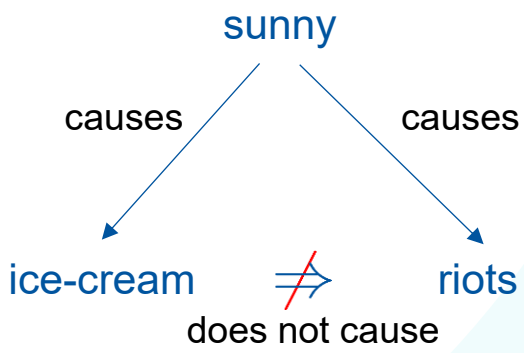
- The hospital in the city of Stolberg has an overall better performance (e.g., lower mortality rate) than the hospital in Aachen. However, for any specific disease, Aachen performs better. This paradox is due to different distributions of diseases (patients with more serious diseases tend to end up in Aachen and not Stolberg).
- Males have higher wages on average, but in any given profession, females earn more on average. This paradox is explained by males going for higher-paid professions.
- Low birth-weight paradox: low birth-weight children born to smoking mothers have a lower infant mortality rate than low-birth-weight children of non-smokers. Smoking is harmful and contributes to low birth weight and higher mortality than normal birth weight. However, other causes of low birth weight are generally more harmful than smoking.

Confounding

- Simpsons paradox is related to confounding
- A (possibly hidden) confounding feature $\mathcal{C}$ (also called "lurking variable") may influence both $\mathcal{A}$ and $\mathcal{B}$, and therefore "blur" $\mathcal{A} \Rightarrow \mathcal{B}$







Summary

- Association rules can be discovered starting from frequent items sets
- Any dataset with instances and feature values can be turned into a multiset of itemsets and used for association rule mining (not just “pure itemsets”)
- Support, confidence, and lift can be used to prune and sort association rules
- Rules should be interpreted carefully (Simpson’s paradox and confounders)

Introduction to Temporal Data

1. **Event Data**
2. Time Series Data
3. Sequence Data

Temporal Data – Discrete Timestamped Events

Time-stamp	f_1	f_2	f_3	f_4	…	f_D
t_1						
t_2						
t_3						
t_4						
t_5						
…						

Every instance happened at a specific time

Temporal Data – Discrete Timestamped Events

Event data

Time-stamp	Case ID	Activity	f_1	f_2	…	f_D
t_1	3	a				
t_2	1	a				
t_3	1	b				
t_4	2	a				
t_5	3	b				
…	…	…				

Case ID is used to group events

Activity identifies the type of event

Event Data

- Timestamp (typically not equal intervals)
- Case ID (maps events to cases)
- Activity (identifies the event type)
- Other features are optional (resource, location, cost, duration, …)

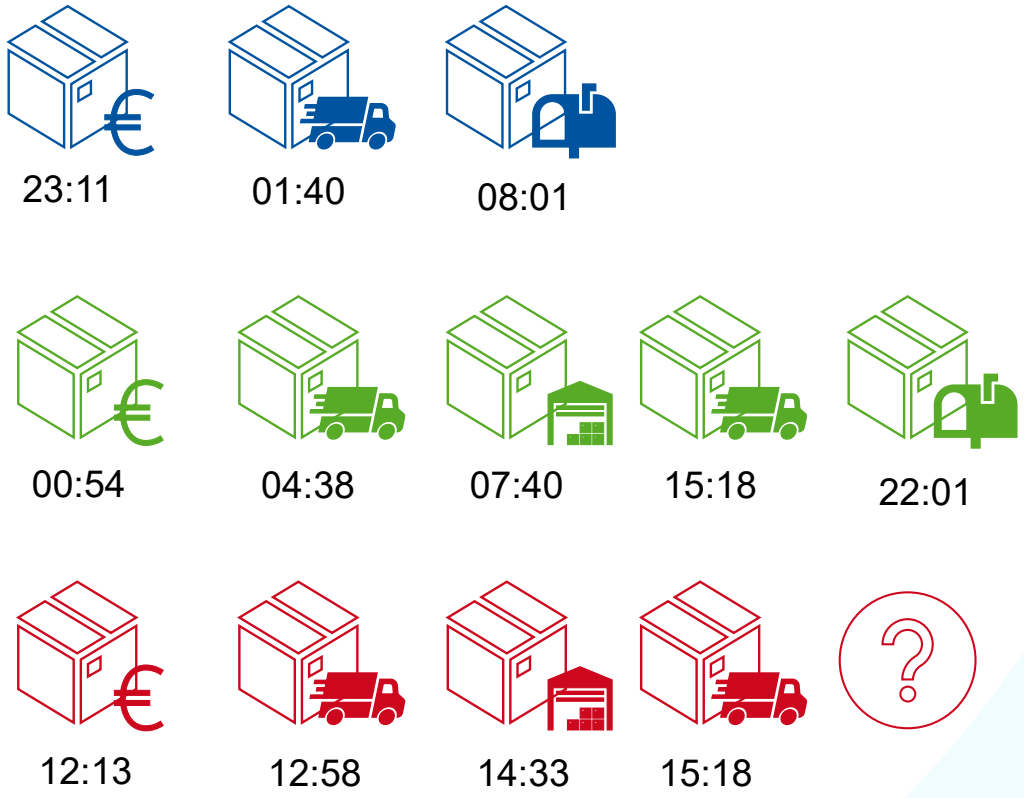
Temporal Data – Discrete Timestamped Events

Event data

Time-stamp	Case ID	Activity	f_1	f_2	…	f_D
t_1	3	a				
t_2	1	a				
t_3	1	b				
t_4	2	a				
t_5	3	b				
…	…	…				

Case ID is used to group events

Activity identifies the type of event



Temporal Data – Discrete Timestamped Events

Event data

Time-stamp	Case ID	Activity	f_1	f_2	…	f_D
t_1	3	a				
t_2	1	a				
t_3	1	b				
t_4	2	a				
t_5	3	b				
…	…	…				

Case ID is used to group events

Activity identifies the type of event

Event Data

- Timestamp (typically not equal intervals)
- Case ID (maps events to cases)
- Activity (identifies the event type)
- Other features are optional (resource, location, cost, duration, …)

$$\text{Case 1: } \langle a, b, \ldots \rangle$$
$$\text{Case 2: } \langle a, \ldots \rangle$$
$$\text{Case 3: } \langle a, b, \ldots \rangle$$

We can abstract from timestamps and optional features to obtain sequences of activities

Temporal Data – Discrete Timestamped Events

Event data

Time-stamp	Case ID	Activity	f_1	f_2	…	f_D
t_1	3	a				
t_2	1	a				
t_3	1	b				
t_4	2	a				
t_5	3	b				
…	…	…				

Case ID is used to group events

Activity identifies the type of event

Event Data

- Timestamp (typically not equal intervals)
- Case ID (maps events to cases)
- Activity (identifies the event type)
- Other features are optional (resource, location, cost, duration, …)

$$\begin{aligned}&\text{Case 1: } \langle a, b, \ldots \rangle \\ &\text{Case 2: } \langle a, \ldots \rangle \\ &\text{Case 3: } \langle a, b, \ldots \rangle\end{aligned} \Rightarrow [\langle a, b, \ldots \rangle^2, \langle a, \ldots \rangle]$$

Event Data – Example 1

Case ID	Activity name	Timestamp	Other features	
Patient ID	**Activity**	**Time**	**Doctor**	**Age**
5611	Blood Test	12:25	Dr. Scott	45
3645	X-Ray	14:34	Dr. House	67
5611	Surgery	15:01	Dr. Scott	45
7891	Blood Test	15:03	Dr. House	24
3645	Radiation Therapy	17:25	Dr. Jenna	81
…	…	…	…	…

$5611 : \langle \text{Blood Test}, \text{Surgery}, \ldots \rangle$

$3645 : \langle \text{X-Ray}, \text{Radiation Therapy}, \ldots \rangle$

$7891 : \langle \text{Blood Test}, \ldots \rangle$

Event Data – Example 2

Case ID	Activity name	Timestamp	Other features		
Order Number	**Activity**	**Time**	**Username**	**Product**	**Quantity**
11152	Register Order	15.12.22 12:25	Carrie192	Iphone 14	1
52690	Ship Order	15.12.22 12:45	Johnny1	Earpods	2
11152	Check Stock	15.12.22 13:01	Carrie192	Iphone 14	1
44891	Handle Payment	30.12.22 18:01	Obelisk	USB-C Charger	3
61238	Cancel Order	11.01.23 17:25	Apex_512	MacBook Air	1
…	…	…	…	…	…

$11152 : \langle \text{Register Order}, \text{Check Stock}, \text{Cancel Order}, \ldots \rangle$

$52690 : \langle \text{Ship Order}, \ldots \rangle$

$44891 : \langle \text{Handle Payment}, \ldots \rangle$

Note: ‘Username’ could also be our Case ID, changing the meaning of data!

Event Data – Example 2

Case ID	Activity name	Timestamp	Other features		
Order Number	**Activity**	**Time**	**Username**	**Product**	**Quantity**
11152	Register Order	15.12.22 12:25	Carrie192	Iphone 14	1
52690	Ship Order	15.12.22 12:45	Johnny1	Earpods	2
11152	Check Stock	15.12.22 13:01	Carrie192	Iphone 14	1
44891	Handle Payment	30.12.22 18:01	Obelisk	USB-C Charger	3
61238	Cancel Order	11.01.23 17:25	Apex_512	MacBook Air	1
…	…	…	…	…	…

$11152 : \langle \text{Register Order}, \text{Check Stock}, \text{Cancel Order}, \ldots \rangle$

$52690 : \langle \text{Ship Order}, \ldots \rangle$

$88721 : \langle \text{Register Order}, \text{Check Stock}, \text{Cancel Order}, \ldots \rangle$

Note: the same sequence can occur multiple times for different cases (multiset of sequences)

Event Data

Basis for Process Mining (separate lecture)

Time-stamp	Case ID	Activity	f_1	f_2	…	f_D
t_1	3	a				
t_2	1	a				
t_3	1	b				
t_4	2	a				
t_5	3	b				
…	…	…				

Case ID is used to group events

Activity identifies the type of event

Process Mining

- Processes generate event data
- Every process execution is a case

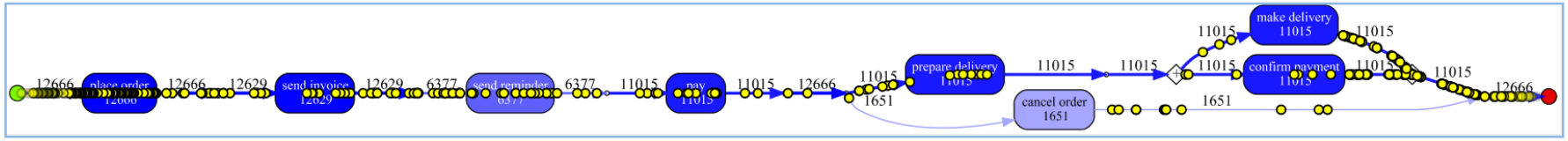
Common Tasks

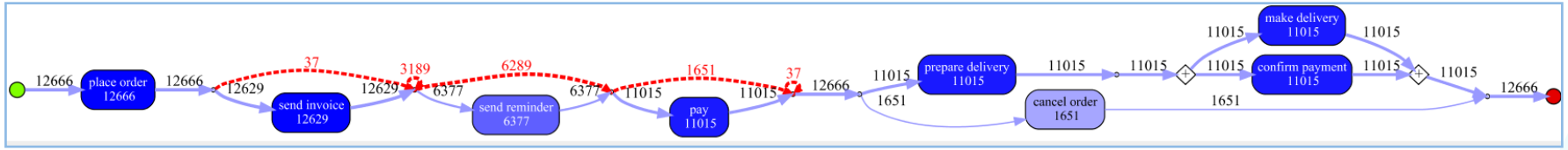
- Discover the process
- Validate the process
- Improve the process

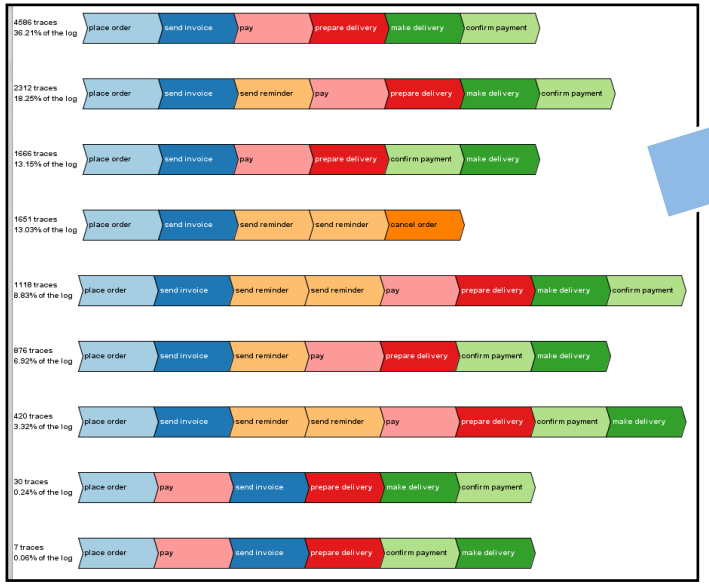
(Separate Lecture)

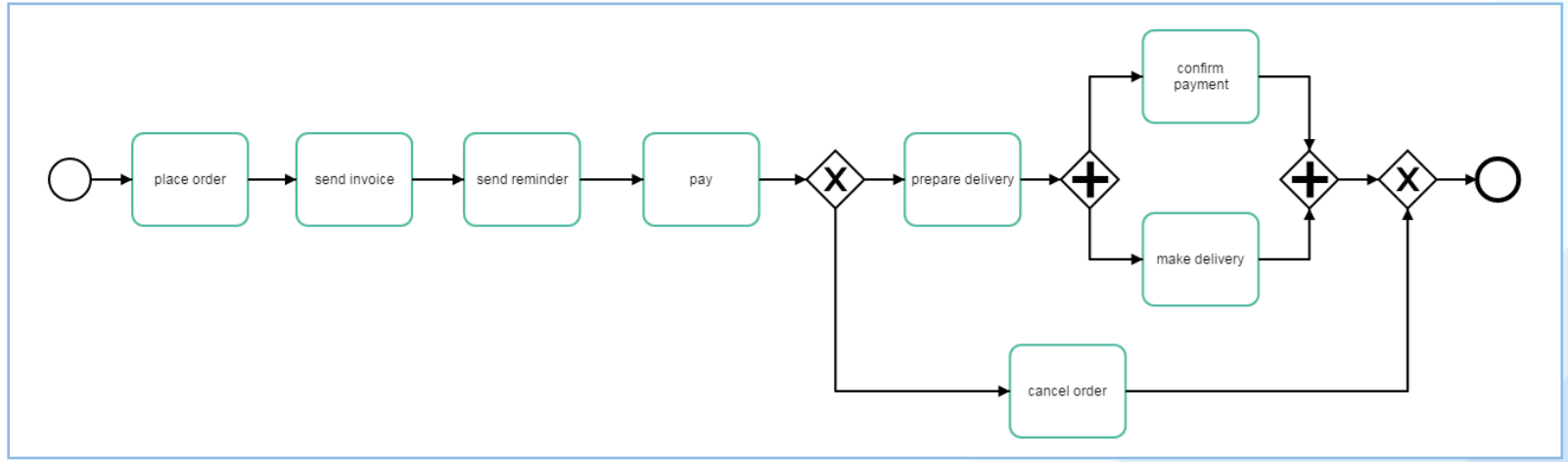
Event Data

Basis for Process Mining (separate lecture)









Temporal Data

1. Event Data
2. **Time Series Data**
3. Sequence Data

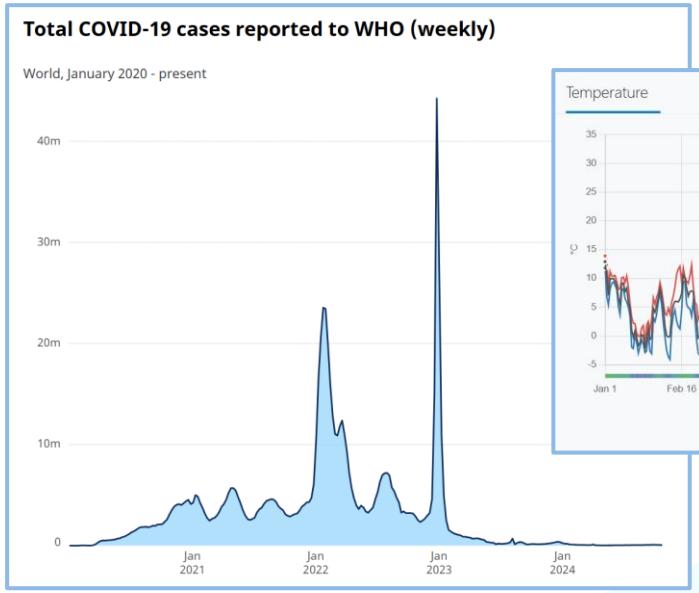
Time Series Data

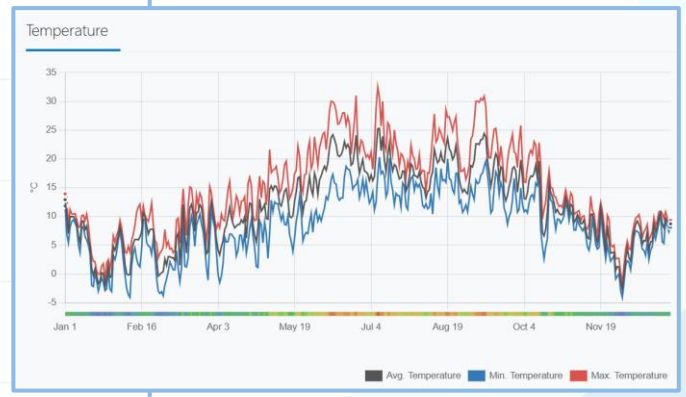
Basis for Time Series Analysis (later lecture)

Time-stamp	f_1	f_2	f_3	f_4	...	f_D
t_1						
t_2						
t_3						
t_4						
t_5						
...						

Intervals are equal

numerical





Temporal Data

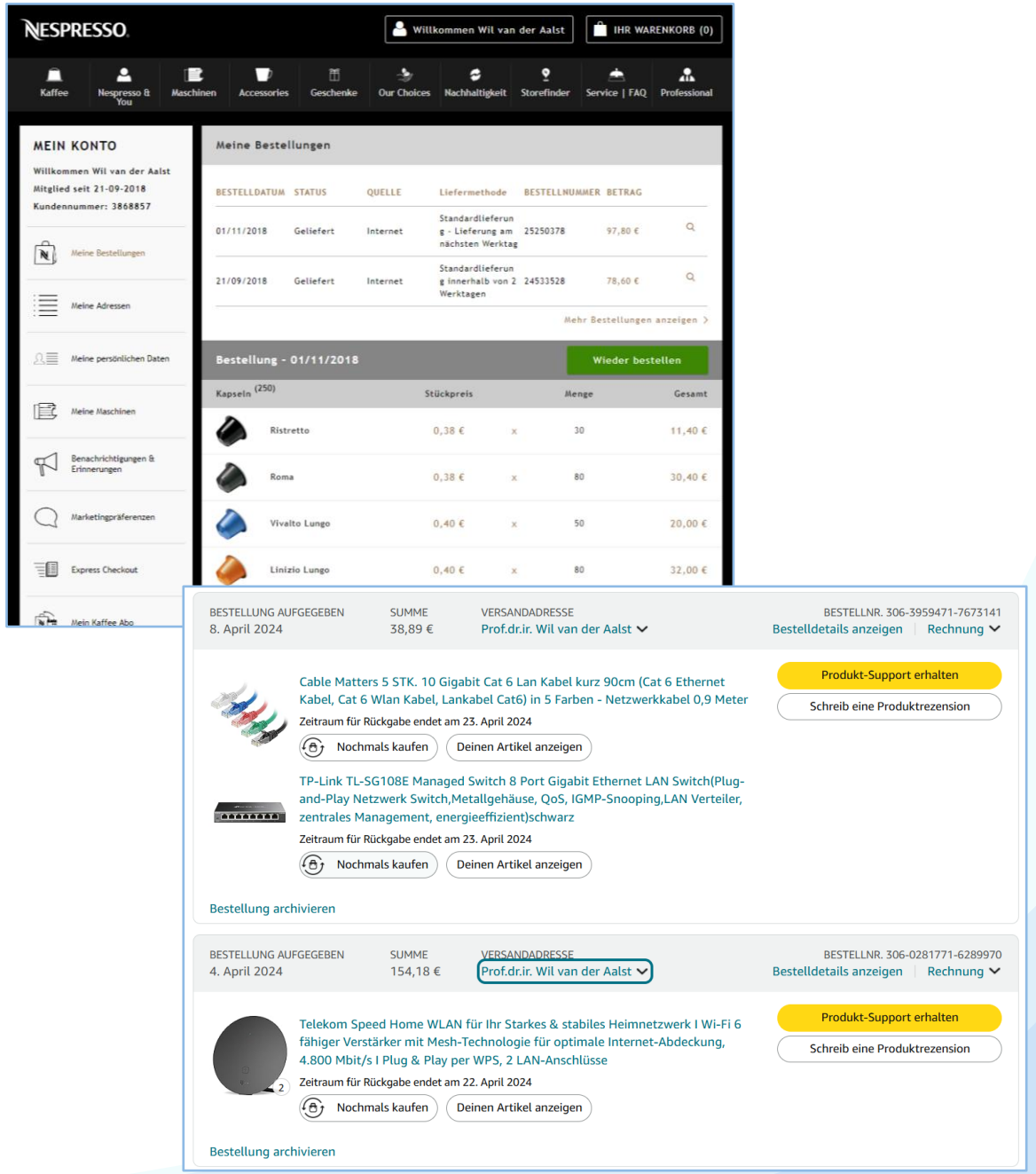
1. Event Data
2. Time Series
3. **Sequence Data**

Sequences of Itemsets

Generalized sequential data

Timestamp	Customer ID	Purchased Itemset
22-07-12	1172	Razor, Shaving Cream
22-07-12	8121	Shampoo
22-08-13	3434	Shampoo
22-09-01	1172	Shaving Cream
...	...	...

$$1172 : \langle \{\text{Razor}, \text{Shaving Cream}\}, \{\text{Shaving Cream}\} \rangle$$
$$8121 : \langle \{\text{Shampoo}\} \rangle$$
$$3434 : \langle \{\text{Shampoo}\} \rangle$$
$$\ldots$$
$$\Rightarrow [\langle \{\text{Razor}, \text{Shaving Cream}\}, \{\text{Shaving Cream}\} \rangle, \langle \{\text{Shampoo}\} \rangle^2, \ldots]$$



Sequences of Itemsets: Used as Input for Sequential Pattern Mining

Customer ID	Purchased Items	Time
1	A	15.12.22 12:25
1	A, B	15.12.22 12:45
2	B	15.12.22 13:01
3	C	30.12.22 18:01
3	A, C, D	11.01.23 17:25
4	B	31.12.22 17:32
…	…	…

Customer ID	Customer Sequence
1	$\langle\{A\},\{A,B\}\rangle$
2	$\langle\{B\}\rangle$
3	$\langle\{C\},\{A,C,D\}\rangle$
4	$\langle\{B\}\rangle$
…	…

Input is a multiset of sequences of itemsets

Frequent Sequence Patterns: Containment

Customer ID	Customer Sequence
1	$\langle\{beer, chips\}, \{wine, chips\}\rangle$
2	$\langle\{beer, chips\}, \{beer\}\rangle$
3	$\langle\{beer\}, \{beer\}, \{wine\}\rangle$
4	$\langle\{beer\}, \{chips\}, \{wine\}\rangle$
5	$\langle\{beer, chips, wine\}\rangle$

$\langle\{beer\}, \{wine\}\rangle$ is contained by 1,3,4

$\langle\{beer, wine\}\rangle$ is contained by 5

$\langle\{beer\}, \{beer\}\rangle$ is contained by 2,3

$\langle\{beer\}\rangle$ is contained by 1,2,3,4,5

$\langle\{beer\}, \{chips, wine\}\rangle$ is contained by 1

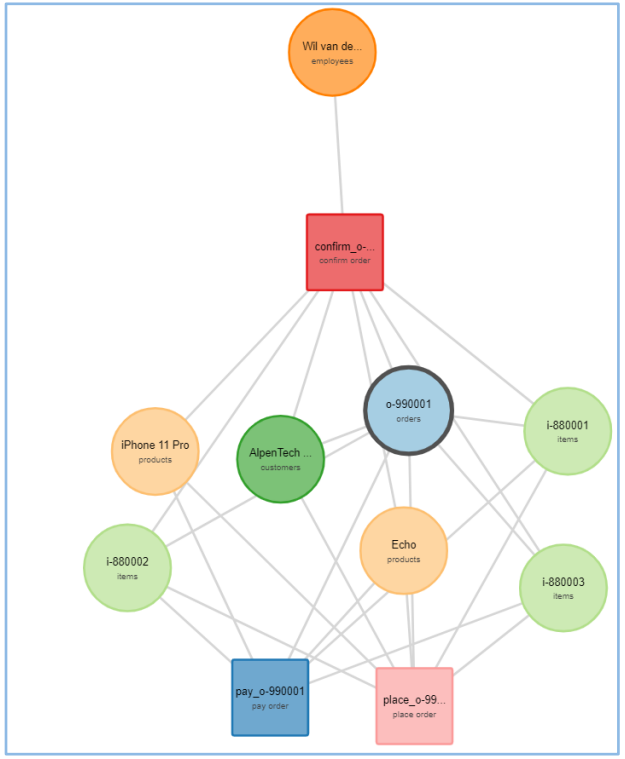
It is possible to apply the apriori principle to determine frequent sequence patterns!

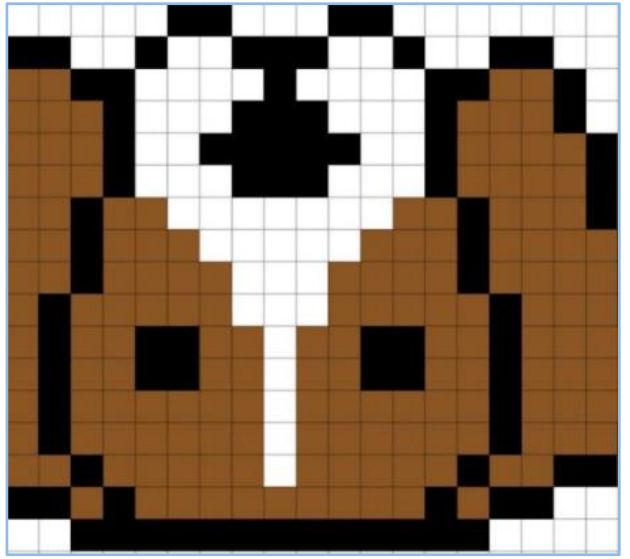
Apriori Principle: Can Be Applied Everywhere

Payment details
Visa ending in 4006
Im Hag 55 , Eschweiler , Nordrhein-Westfalen 52249 Germany

Description	Quantity	Shipping	Product Price	Tax Rate	Tax Amount	Product Total
IEEE Computational Intelligence Society Membership **Included** - Games, IEEE Transactions on Format : Electronic	1	N/A	$29.00			$29.00
IEEE Standards Association Individual Membership	1	N/A	$68.00			$68.00
IEEE Membership **Included** - Spectrum, IEEE Format : Print - Potentials Magazine, IEEE Format : Electronic	1	N/A	$188.00			$188.00
IEEE Computer Society Membership **Included** - Computer Magazine, IEEE Format : Electronic	1	N/A	$60.00			$60.00

Net Amount:	$345.00
Shipping and Handling:	$0.00
Tax:	$0.00





A pattern cannot be frequent if it contains an infrequent smaller pattern.

Conclusion

- Association Rules: Challenges and Applications
- Support, Confidence, Lift
- Using features values as items and instances as itemsets
- Simpson’s Paradox
- Temporal Data: Event Data, Time Series, Sequence Data
- Apriori: A universal principle