Machine Learning

Learning from Data

Prof. Dr. – Ing. Morris Riedel
Associated Professor
School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland
Research Group Leader, Juelich Supercomputing Centre, Forschungszentrum Juelich, Germany

@Morris Riedel @MorrisRiedel @MorrisRiedel

LECTURE 1

Association Rule Mining – Apriori Example

April 24, 2020
Webinar, ON4OFF Project Meeting

ON 4 OFF

UNIVERSITY OF ICELAND
SCHOOL OF ENGINEERING AND NATURAL SCIENCES
FACULTY OF INDUSTRIAL ENGINEERING,
MECHANICAL ENGINEERING AND COMPUTER SCIENCE

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

DEEP Projects

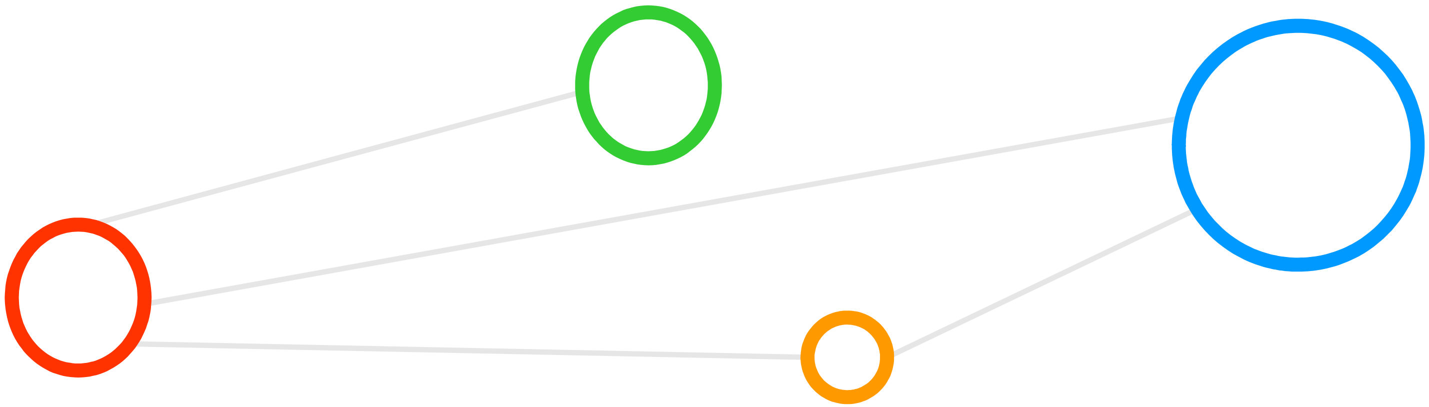
HELMHOLTZAI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

Outline

- Introduction to Association Rule Mining
  - Machine Learning Prerequisited & Methodology
  - Role of Datasets & Famous Dataset Example
  - Challenges & Limits in Reality
  - Support, Confidence, Lift, Leverage & Conviction Algorithm Options
  - Relevant Rules & Practical Approach

- Practical Apriori Example ‘Pieper Store‘
  - Apriori Algorithm & Association Rules
  - Using Python Mlxtend Library
  - Jupyter Notebook Demonstration
  - Fake Data(!) created from Supermarket Data
  - Examples in Context

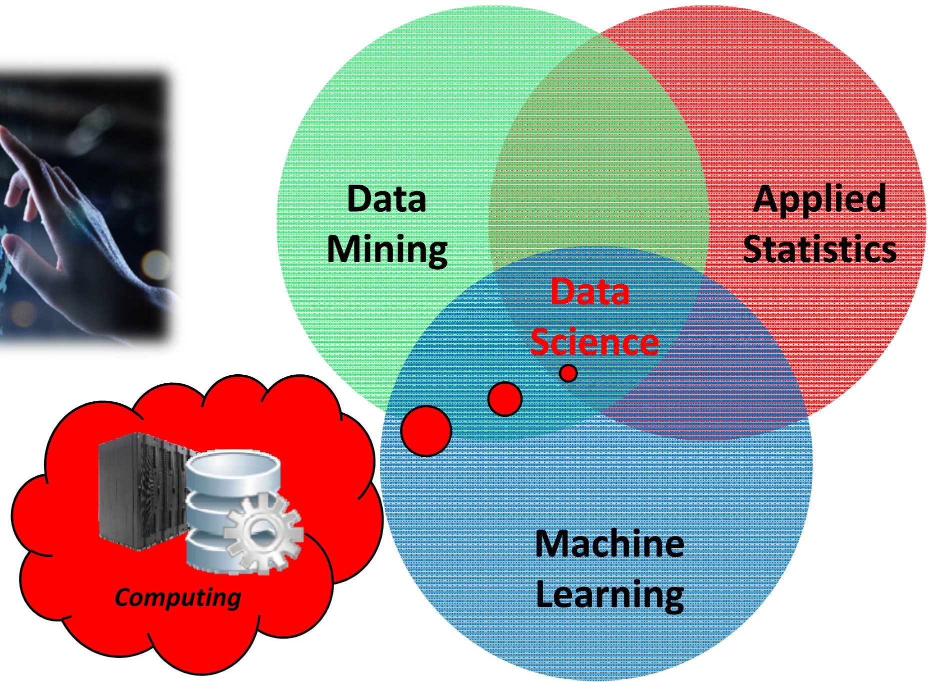
Association Rule Mining



Machine Learning Prerequisites & Challenges

1. Some pattern exists
2. No exact mathematical formula
3. Data exists

- Idea 'Learning from Data'
  - Shared with a wide variety of other disciplines
  - E.g. signal processing, data mining, etc.
  - Association Rule Mining needs data!
- Challenges
  - Data is often complex
  - Learning from data requires processing time → Clouds



- **Machine learning is a very broad subject and goes from very abstract theory to extreme practice ('rules of thumb')**
- **Training machine learning models needs processing time**

Association Rule Mining – Methodology

- Methodology
  - Sometimes referred to as simply 'Association Rules'
  - Used to discover unknown relationships hidden in datasets
  - Rules refer to a set of identified frequent itemsets that represent the uncovered relationships in datasets
  - Identify rules that will predict the occurence of one or more items based on the occurrences of other items in the dataset

- Approach
  - Unsupervised machine learning method
  - No direct guiding output data is given to find the patterns
  - Several algorithms exist to perform association rule mining (e.g., Apriori, FP Growth, MAFIA, etc.)

(this lecture will focus on Apriori with a simple demonstration)

[1] Big Data Tips, Association Rules

Association Rule Mining – Role of Datasets

- Commercial Environments
  - Large quantities of data are accumulated in databases from day-to-day operations
  - Lays the foundation for mining association rules: no data – no association rule mining!
- Retail Example
  - Customer purchase data are collected on a daily basis at the checkout counters of city stores or when shopping at online stores
  - Accumulated data items are often market basket transactions
- Motivation to Collect and Analyze Data
  - Managers of stores are interested in analyzing the collected data in order to learn the purchasing behaviour of customers
  - Enables a large variety of business-related applications based on the identified rules in the data (to be reviewed from store managers!)

(careful: can be different shopping patterns offline vs. online)

[1] Big Data Tips, Association Rules

Association Rule Mining – Famous Dataset Example & Challenges

- Famous Example in Retail
  - Illustrating a rule based on a strong relationship between the sale of Diapers and the sale of Beer
  - Many customers who buy Diapers also buy Beer
  - Investigating the transactions to find those frequent itemsets seems to be easy
- Challenges
  - In real datasets millions or billions of transactions are searched
  - Transaction search across 100000 of different items that may identify 1000 of rules
- Algorithms Benefit
  - Automation of the process using association rule mining algorithms.
  - Rules help to identify new opportunities and ways for cross-selling products to customers

ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
...	...

market basket transactions

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule

[1] Big Data Tips, Association Rules

Association Rule Mining – Famous Dataset Example & Limits

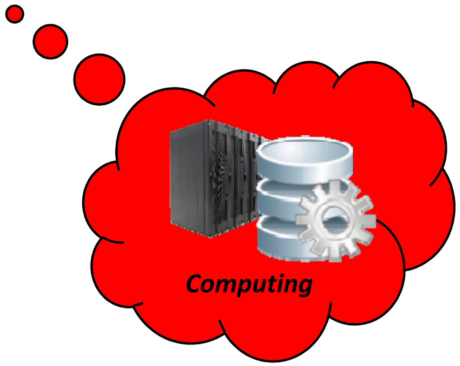
- Frequent Itemsets Limits
  - Very simple method
  - E.g., no personalized shopping behaviour
  - Ignores other relevant attributes
  - E.g., quantity of items sold, the price, or even a specific brand of items
- Computational Complexity
  - Mining process to discover unknown patterns in large transaction datasets is computationally expensive
  - Size of the millions or billions of the transaction dataset when performing mining within memory can be tricky
  - Limits with sub-sampling of dataset items since this may increase the risk to overlook frequent itemset patterns
  - E.g., use smart out-of-cpu strategies to work on the big datasets as a whole to identify frequent itemsets and/or use High Performance Computing (HPC)

ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
…	…

market basket transactions

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule



Association Rule Mining – Reality & Configuration Options

- Facing Reality in Retail Shopping
  - Some uncovered patterns are simply not true
  - Reasoning: they have happen by chance!
- Approaches
  - Manual post-processing or even application domain knowledge to find true rules out of the undiscovered patterns
  - Important is to find actionable insights from rules that can be used to change a product portfolio, store setup, or customer relations
- Configuration Options
  - In order to provide a more clear set of rules
  - E.g. support, confidence

(actionable insight:
change a product portfolio)

[2] Image: Luehrmann.de

(actionable insight:
change a store setup)

(actionable insight:
change customer relations)

Association Rule Mining – Understanding Support Option

- Practical Example
  - Text analysis
- Example 'Dog'
  - Appears in 7 baskets
  - Not in (5)
  - Support is 7
- Example 'Cat'
  - Appears in 6 baskets
  - Not (4) and (8)
  - Support is 6
- Example 'support threshold'
  - Pick s = 3 results in '5 frequent singleton itemsets'

('Example of 8 baskets consisting of items as words')

1. {Cat, and, dog, bites}
2. {Yahoo, news, claims, a, cat, mated, with, a, dog, and, produced, viable, offspring}
3. {Cat, killer, likely, is, a, big, dog}
4. {Professional, free, advice, on, dog, training, puppy, training}
5. {Cat, and, kitten, training, and, behavior}
6. {Dog, &, Cat, provides, dog, training, in, Eugene, Oregon}
7. {"Dog, and, cat", is, a, slang, term, used, by, police, officers, for, a, male–female, relationship}
8. {Shop, for, your, show, dog, grooming, and, pet, supplies}

(Basket #4 has 2x training: as baskets are modelled as sets this can be ignored)

- **Careful: MLxtend implementation min_support threshold is set to 0.5 (50%), a frequent itemset is defined as a set of items that occur together in at least 50% of all transactions in the database**

[4] mlxtend lib, Apriori

[5] Mining Large Datasets

Association Rule Mining – Understanding Support Option

- Metrics
  - Evaluating association rules and setting selection thresholds
  - Given a rule "A -> C“
  - *A* stands for antecedent and *C* stands for consequent
- Support – Revisited
  - Defined for itemsets, not assocication rules
- Example MLxtend lib
  - Table produced by the association rule mining algorithm contains three different support metrics
  - 'antecedent support' computes the proportion of transactions that contain the antecedent A
  - 'consequent support' computes the support for the itemset of the consequent C
  - 'support' metric then computes the support of the combined itemset A U C

ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
...	...

market basket transactions

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule

$$\text{support}(A \rightarrow C) = \text{support}(A \cup C), \quad \text{range: } [0, 1]$$

Typically, support is used to measure the abundance or frequency (often interpreted as significance or importance) of an itemset in a database. We refer to an itemset as a "frequent itemset" if you support is larger than a specified minimum-support threshold. Note that in general, due to the *downward closure* property, all subsets of a frequent itemset are also frequent.

[6] MLxtend Lib, Association Rules

Association Rule Mining – Understanding Confidence Option

- Approach
  - Finding Association Rules with high confidence important
  - Identifying useful association rules is not much harder than finding frequent itemsets
- Practical insights
  - Looking for association rules $I \rightarrow j$ that apply to a reasonable fraction of the baskets: the support of I must be reasonably high
  - Marketing in brick-and-mortar stores w.r.t support: 'reasonably high' is often around 1% of the baskets
  - Confidence of the rule should be reasonably high, e.g. 50% (otherwise the rule has little practical effect)
  - As a result, the set $I \cup \{j\}$ will also have fairly high support.

ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
...	...

market basket transactions

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule

(Finding within all the association rules those that have both high support and high confidence is possible)

[5] Mining Large Datasets

Association Rule Mining – Understanding Confidence & Lift Option

- Example MLxtend lib – Confidence
  - Confidence of a rule A->C is the probability of seeing the consequent in a transaction given that it also contains the antecedent
  - Metric is not symmetric or directed
  - E.g, the confidence for A->C is different than the confidence for C->A
  - Confidence is 1 (maximal) for a rule A->C if the consequent and antecedent always occur together
- Example MLxtend lib – Lift
  - Lift metric is commonly used to measure how much more often the antecedent and consequent of a rule A->C occur together than we would expect if they were statistically independent
  - E.g., if A and C are independent, the Lift score will be exactly 1

ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
…	…

market basket transactions

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule

$$\text{confidence}(A \rightarrow C) = \frac{\text{support}(A \rightarrow C)}{\text{support}(A)}, \quad \text{range: } [0, 1]$$

$$\text{lift}(A \rightarrow C) = \frac{\text{confidence}(A \rightarrow C)}{\text{support}(C)}, \quad \text{range: } [0, \infty]$$

[6] MLxtend Lib, Association Rules

Association Rule Mining – Understanding Leverage & Conviction Option

- Example MLxtend lib – Leverage
  - Leverage computes the difference between the observed frequency of A and C appearing together and the frequency that would be expected if A and C were independent
  - An leverage value of 0 indicates independence

ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
…	…

market basket transactions

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule

$$\text{levarage}(A \to C) = \text{support}(A \to C) - \text{support}(A) \times \text{support}(C), \quad \text{range: } [-1, 1]$$

- Example MLxtend lib – Conviction
  - High conviction value means that the consequent is highly depending on the antecedent
  - E.g., in the case of a perfect confidence score, the denominator becomes 0 (due to 1 - 1) for which the conviction score is defined as 'inf'
  - Similar to lift, if items are independent, the conviction is 1

$$\text{conviction}(A \to C) = \frac{1 - \text{support}(C)}{1 - \text{confidence}(A \to C)}, \quad \text{range: } [0, \infty]$$

[6] MLxtend Lib, Association Rules

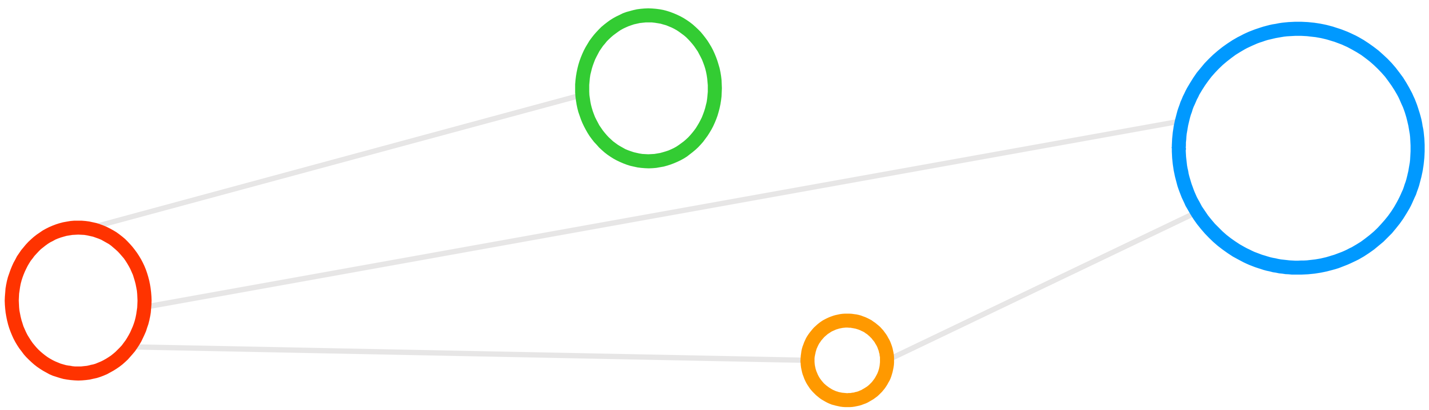
Relevant Rules & Practical Approach

- Relevant Rules
  - Assumed that there are not too many frequent itemsets
  - Not many candidates for high-support, high-confidence association rules.
- ‘less is more‘
  - Each one found from above must be acted upon
  - If store manager receives a million association rules that meet thresholds (e.g. for support and confidence) they cannot read them or act on them
- Practical Approach
  - Normal to adjust the support threshold s
  - Provides not ‘too many frequent itemsets’
  - Implement and review with store manager step-wise to measure results better

(Generating ‘big data‘: Major chains might sell 100.000 different items and collect data about millions of market baskets)

- **The massive amount of data and derived rules requires an automated algorithm-based approach**
- **The most commonly used algorithm (and its variants) for association rules mining is called Apriori**

Practical Apriori Example ‘Pieper Store’



Jupyter Notebook Demonstration

Python Script using MLxtend Library

```
In [ ]: import pandas as pd
        import numpy as np
        from mlxtend.frequent_patterns import apriori, association_rules
        import matplotlib.pyplot as plt
```

```
In [ ]: df = pd.read_csv('./data/retail_dataset-pieper-fake.csv', sep=',')
        df.head()
```

```
In [ ]: items = (df['0'].unique())
        items
```

```
In [ ]: # transform dataset to one-hot encoded dataset
        encoded_items = []
        def onehotencoding():
            for index, row in df.iterrows():
                present = {}
                uncommons = list(set(items) - set(row))
                commons = list(set(items).intersection(row))
                for uc in uncommons:
                    present[uc] = 0
                for com in commons:
                    present[com] = 1
                encoded_items.append(present)
        onehotencoding()
        ohe_df = pd.DataFrame(encoded_items)
        print(ohe_df)
```

```
In [ ]: # running apriori algorithm
        freq_items = apriori(ohe_df, min_support=0.2, use_colnames=True, verbose=1)
        freq_items.head(7)
```

```
In [ ]: # using frequent itemsets from apriori for association rules
        rules = association_rules(freq_items, metric="confidence", min_threshold=0.6)
        rules.head(7)
```

changed from [7] MLxtend tutorial

Pieper Store Example – Fake Data

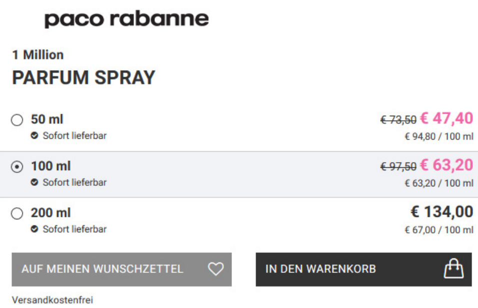
[3] Pieper.de Duefte

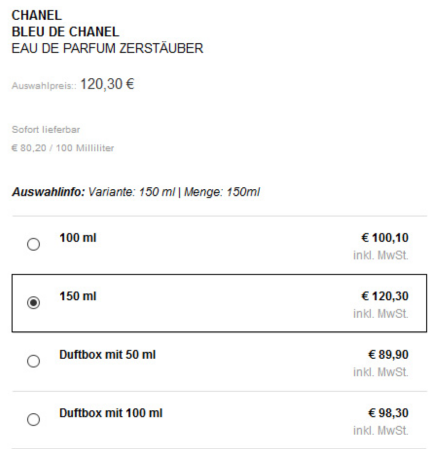
- Retail Data
  - Original

```
0,1,2,3,4,5,6
Bread,Wine,Eggs,Meat,Cheese,Pencil,Diaper
Bread,Cheese,Meat,Diaper,Wine,Milk,Pencil
Cheese,Meat,Eggs,Milk,Wine,,
Cheese,Meat,Eggs,Milk,Wine,,
Meat,Pencil,Wine,,,,
Eggs,Bread,Wine,Pencil,Milk,Diaper,Bagel
Wine,Pencil,Eggs,Cheese,,,
Bagel,Bread,Milk,Pencil,Diaper,,
Bread,Diaper,Cheese,Milk,Wine,Eggs,
Bagel,Wine,Diaper,Meat,Pencil,Eggs,Cheese
Cheese,Meat,Eggs,Milk,Wine,,
Bagel,Eggs,Meat,Bread,Diaper,Wine,Milk
Bread,Diaper,Pencil,Bagel,Meat,,
Bagel,Cheese,Milk,Meat,,,
Bread,,,,,,
Pencil,Diaper,Bagel,,,,
Meat,Bagel,Bread,,,,
Bread,Bagel,Milk,,,,
Diaper,,,,,,
Bagel,Cheese,Meat,Bread,Diaper,Eggs,
Meat,Pencil,Cheese,Bread,,,
Cheese,Eggs,Wine,Bread,Milk,Pencil,Meat
Eggs,Bagel,Cheese,Meat,Diaper,,
```

(for simplicity we focus here on the brand names and not on the detailed description, sizes in ml, or different prices)

```
0,1,2,3,4,5,6
HugoBoss,Chanel,Rituals,Armani,Biotherm,Lancome,Dior
HugoBoss,Biotherm,Armani,Dior,Chanel,Clinique,Lancome
Biotherm,Armani,Rituals,Clinique,Chanel,,
Biotherm,Armani,Rituals,Clinique,Chanel,,
Armani,Lancome,Chanel,,,,
Rituals,HugoBoss,Chanel,Lancome,Clinique,Dior,Clinique
Chanel,Lancome,Rituals,Biotherm,,,
Clinique,HugoBoss,Clinique,Lancome,Dior,,
HugoBoss,Dior,Biotherm,Clinique,Chanel,Rituals,
Clinique,Chanel,Dior,Armani,Lancome,Rituals,Biotherm
Biotherm,Armani,Rituals,Clinique,Chanel,,
Clinique,Rituals,Armani,HugoBoss,Dior,Chanel,Clinique
HugoBoss,Dior,Lancome,Clinique,Armani,,
Clinique,Biotherm,Clinique,Armani,,,
HugoBoss,,,,,,
Lancome,Dior,Clinique,,,,
Armani,Clinique,HugoBoss,,,,
HugoBoss,Clinique,Clinique,,,,
Dior,,,,,,
Clinique,Biotherm,Armani,HugoBoss,Dior,Rituals,
Armani,Lancome,Biotherm,HugoBoss,,,
Biotherm,Rituals,Chanel,HugoBoss,Clinique,Lancome,Armani
Rituals,Clinique,Biotherm,Armani,Dior,,
```





Association Rule Mining – Understanding Support Option – Revisited

- Example MLxtend lib
  - Table produced by the association rule mining algorithm contains three different support metrics
  - 'antecedent support' computes the proportion of transactions that contain the antecedent A
  - 'consequent support' computes the support for the itemset of the consequent C
  - 'support' metric then computes the support of the combined itemset A U C

$$\text{support}(A \rightarrow C) = \text{support}(A \cup C), \quad \text{range: } [0, 1]$$

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Dior)	(Clinique)	0.406349	0.701587	0.260317	0.640625	0.913108	-0.024772	0.830366
1	(Biotherm)	(Clinique)	0.501587	0.701587	0.409524	0.816456	1.163726	0.057617	1.625835
2	(HugoBoss)	(Clinique)	0.504762	0.701587	0.387302	0.767296	1.093657	0.033167	1.282368
3	(Chanel)	(Clinique)	0.438095	0.701587	0.301587	0.688406	0.981212	-0.005775	0.957697
4	(Rituals)	(Clinique)	0.438095	0.701587	0.333333	0.760870	1.084497	0.025971	1.247908
5	(Armani)	(Clinique)	0.476190	0.701587	0.358730	0.753333	1.073756	0.024641	1.209781
6	(Lancome)	(Clinique)	0.361905	0.701587	0.260317	0.719298	1.025244	0.006410	1.063095

[6] MLxtend Lib, Association Rules

Association Rule Mining – Understanding Confidence & Lift Option – Revisited

- Example MLxtend lib – Confidence
  - Confidence of a rule A->C is the probability of seeing the consequent in a transaction given that it also contains the antecedent
  - Metric is not symmetric or directed
  - E.g, the confidence for A->C is different than the confidence for C->A
  - Confidence is 1 (maximal) for a rule A->C if the consequent and antecedent always occur together

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Dior)	(Clinique)	0.406349	0.701587	0.260317	0.640625	0.913108	-0.024772	0.830366
1	(Biotherm)	(Clinique)	0.501587	0.701587	0.409524	0.816456	1.163726	0.057617	1.625835
2	(HugoBoss)	(Clinique)	0.504762	0.701587	0.387302	0.767296	1.093657	0.033167	1.282368
3	(Chanel)	(Clinique)	0.438095	0.701587	0.301587	0.688406	0.981212	-0.005775	0.957697
4	(Rituals)	(Clinique)	0.438095	0.701587	0.333333	0.760870	1.084497	0.025971	1.247908
5	(Armani)	(Clinique)	0.476190	0.701587	0.358730	0.753333	1.073756	0.024641	1.209781
6	(Lancome)	(Clinique)	0.361905	0.701587	0.260317	0.719298	1.025244	0.006410	1.063095

$$\text{confidence}(A \rightarrow C) = \frac{\text{support}(A \rightarrow C)}{\text{support}(A)}, \quad \text{range: } [0, 1]$$

- Example MLxtend lib – Lift
  - Lift metric is commonly used to measure how much more often the antecedent and consequent of a rule A->C occur together than we would expect if they were statistically independent
  - E.g., if A and C are independent, the Lift score will be exactly 1

$$\text{lift}(A \rightarrow C) = \frac{\text{confidence}(A \rightarrow C)}{\text{support}(C)}, \quad \text{range: } [0, \infty]$$

[6] MLxtend Lib, Association Rules

Association Rule Mining – Understanding Leverage & Conviction Option

- Example MLxtend lib – Leverage
  - Leverage computes the difference between the observed frequency of A and C appearing together and the frequency that would be expected if A and C were independent
  - An leverage value of 0 indicates independence

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Dior)	(Clinique)	0.406349	0.701587	0.260317	0.640625	0.913108	-0.024772	0.830366
1	(Biotherm)	(Clinique)	0.501587	0.701587	0.409524	0.816456	1.163726	0.057617	1.625835
2	(HugoBoss)	(Clinique)	0.504762	0.701587	0.387302	0.767296	1.093657	0.033167	1.282368
3	(Chanel)	(Clinique)	0.438095	0.701587	0.301587	0.688406	0.981212	-0.005775	0.957697
4	(Rituals)	(Clinique)	0.438095	0.701587	0.333333	0.760870	1.084497	0.025971	1.247908
5	(Armani)	(Clinique)	0.476190	0.701587	0.358730	0.753333	1.073756	0.024641	1.209781
6	(Lancome)	(Clinique)	0.361905	0.701587	0.260317	0.719298	1.025244	0.006410	1.063095

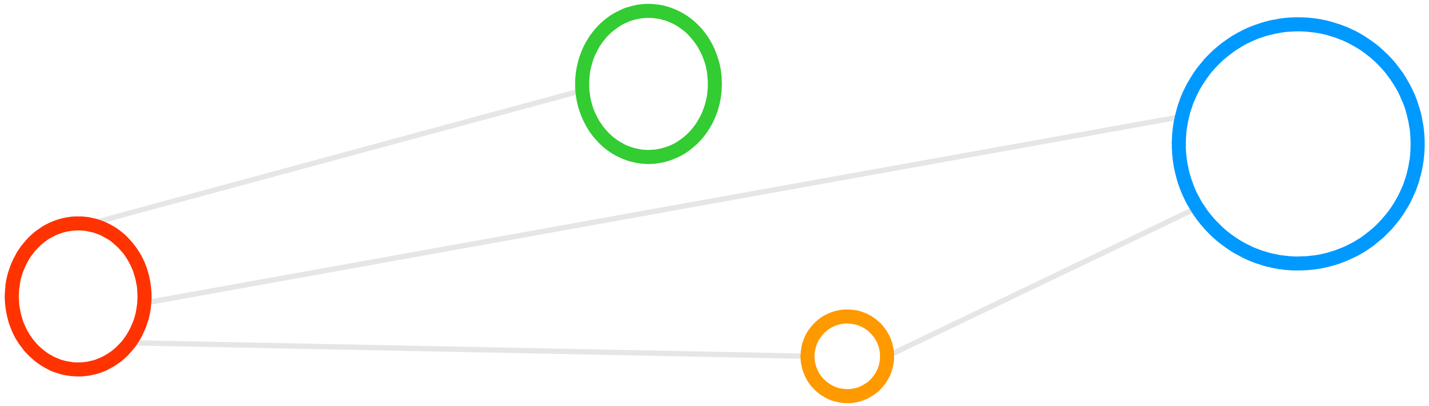
$$\text{levarage}(A \rightarrow C) = \text{support}(A \rightarrow C) - \text{support}(A) \times \text{support}(C), \quad \text{range: } [-1, 1]$$

- Example MLxtend lib – Conviction
  - High conviction value means that the consequent is highly depending on the antecedent
  - E.g., in the case of a perfect confidence score, the denominator becomes 0 (due to 1 - 1) for which the conviction score is defined as 'inf'
  - Similar to lift, if items are independent, the conviction is 1

$$\text{conviction}(A \rightarrow C) = \frac{1 - \text{support}(C)}{1 - \text{confidence}(A \rightarrow C)}, \quad \text{range: } [0, \infty]$$

[6] Mlxtend Lib, Association Rules

Lecture Bibliography



Lecture Bibliography

- [1] Big Data Tips, Association Rules, Online: http://www.big-data.tips/association-rules
- [2] Luehrmann.de, ‘Parfuemerie Pieper bezieht markante Eckimmobilie‘, Online: https://www.luehrmann.de/de/city-life/erfolgsraume/beitrag/parfumerie-pieper-bezieht-markante-eckimmobilie/
- [3] Pieper.de Duefte, Online: https://www.pieper.de/duefte/
- [4] MLxtend Library, Apriori: http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/
- [5] Mining of Massive Datasets, Anand Rajaraman, Jure Leskovec, Jeffrey D. Ullman, Cambridge University Press, ISBN 1107015359, ~326 pages, English, 2011
- [6] MLxtend Library, Association Rules: http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/
- [7] Tutorial, Association Analysis in Python, Online: https://medium.com/analytics-vidhya/association-analysis-in-python-2b955d0180c

services
measurement
within
fusion
new
funding
Policy-based
concepts
forms
cross-disciplinary
resources
device
Centre
Analysis
climate
Computer
expertise
year
like
Computational
Test
E.g
Computing
disciplines
results
Approaches
construction
Fusion
Enable
Cross-Disciplinary
EU
methods
High
big
nodes
PB
community
Environment
Modelling
different
Science
centers
Key
scientific
important
files
model
international
SMAQ
engineering
References
etc
DKRZ
Use
storage
basis
Infrastructure
simulations
ENES
technologies
formats
brain
increasing
via
EUDAT
physical
IPCC
Big
Project
Data
MapReduce
create
processing
Common
TB
Methods
computing
Using
European
databases
codes
GB
Supercomputing
Work
images
analysis
Large
Summary
data-intensive
Service
Scientific
levels
often
DLCL
stored
Resources
analyze
external
use
JSC
computational
Level
Research
Matters
Concrete
performance
research
e.g
science
HPC
large-scale
Juelich
system
datasets
acquisition
compute
Landscape
management
System
Climate
modelling
AR5
Hardware
distributed
cores
access
hundreds
Nutshell
Services
DLCLs
Understanding
structures
Technologies
Earth
Structure
current
network
Germany
Simulation
directory
project
General
Provide
NASA
Energy
systems
day
Processing
Health