

Data Mining Tutorial

Evaluation, Apriori

Erich Schubert, Arthur Zimek

Ludwig-Maximilians-Universität München

2013-07-12 — KDD Übung

Nicht ganz einfach:

- ▶ Unbalanciertes Problem
- ▶ Scores oft nicht aussagekräftig
- ▶ Labels i.d.R. nicht vollständig (weitere, "uninteressante" Ausreißer)
- ▶ i.d.R. dennoch "supervised" Evaluation
- ▶ besser eigentlich: Validieren durch Experten!

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Precision@2:

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Precision@2: jeweils 1/2. Recall: 1/2. F-Measure: 1/2.

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Precision@2: jeweils 1/2. Recall: 1/2. F-Measure: 1/2.
 Sehr grobes Maß! 0/2, 1/2 oder 2/2 richtig!

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$AveP(S_1, 4) := \frac{1}{4} (1/1 +$$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$AveP(S_1, 4) := \frac{1}{4} (1/1 + 1/2 +$$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$AveP(S_1, 4) := \frac{1}{4} (1/1 + 1/2 + 1/3 +$$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$\text{AveP}(S_1, 4) := \frac{1}{4} (1/1 + 1/2 + 1/3 + 2/4)$$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$\text{AveP}(S_1, 4) := \frac{1}{4} (1/1 + 1/2 + 1/3 + 2/4) = 7/12$$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$AveP(S_1, 4) := \frac{1}{4} (1/1 + 1/2 + 1/3 + 2/4) = 7/12$$

$$AveP(S_2, 4) := \frac{1}{4} ($$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$\text{AveP}(S_1, 4) := \frac{1}{4} (1/1 + 1/2 + 1/3 + 2/4) = 7/12$$

$$\text{AveP}(S_2, 4) := \frac{1}{4} (0/1 +$$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$AveP(S_1, 4) := \frac{1}{4} (1/1 + 1/2 + 1/3 + 2/4) = 7/12$$

$$AveP(S_2, 4) := \frac{1}{4} (0/1 + 1/2 +$$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$AveP(S_1, 4) := \frac{1}{4} (1/1 + 1/2 + 1/3 + 2/4) = 7/12$$

$$AveP(S_2, 4) := \frac{1}{4} (0/1 + 1/2 + 2/3 +$$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$\text{AveP}(S_1, 4) := \frac{1}{4} (1/1 + 1/2 + 1/3 + 2/4) = 7/12$$

$$\text{AveP}(S_2, 4) := \frac{1}{4} (0/1 + 1/2 + 2/3 + 2/4)$$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

Average Precision für $k = [1 \dots 4]$:

$$AveP(S_1, 4) := \frac{1}{4} (1/1 + 1/2 + 1/3 + 2/4) = 7/12$$

$$AveP(S_2, 4) := \frac{1}{4} (0/1 + 1/2 + 2/3 + 2/4) = 5/12$$

ROC-Kurven: False-Positive-Rate vs. True-Positive-Rate für *alle* k ! Implizit gewichtet nach Klassengröße, da x -Achse: $|I|$ Stufen, y -Achse: $|O|$ Stufen!

Aufgabe 11-1

Aufgabe 11-2

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

ROC-Kurven: False-Positive-Rate vs. True-Positive-Rate für *alle* k ! Implizit gewichtet nach Klassengröße, da x -Achse: $|I|$ Stufen, y -Achse: $|O|$ Stufen!

Aufgabe 11-1

Aufgabe 11-2

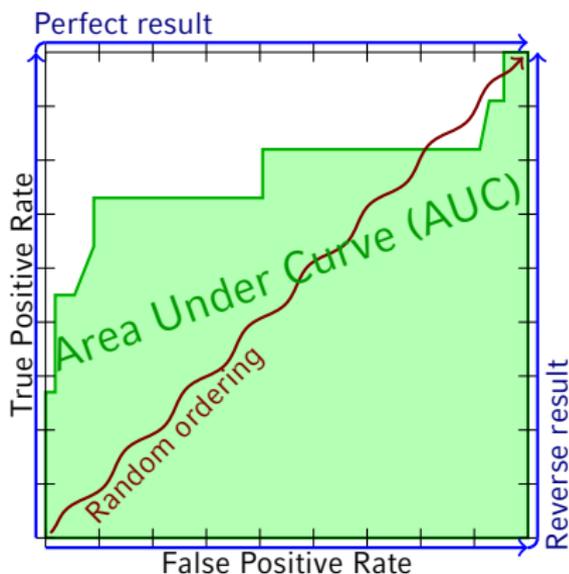
C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

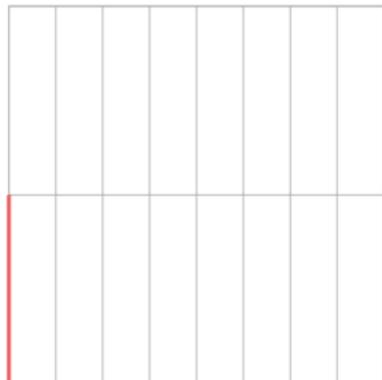


Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10

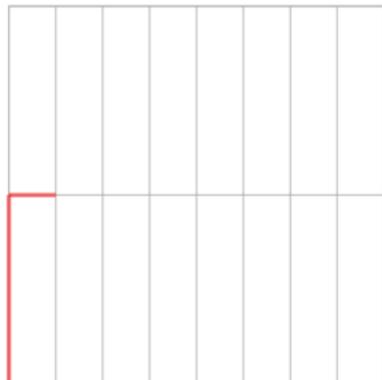
Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



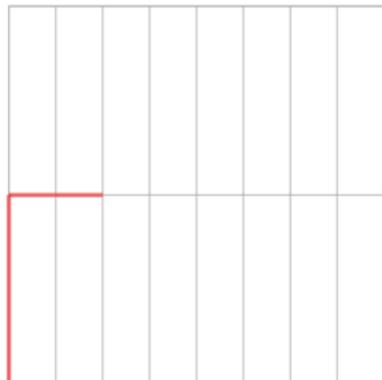
Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



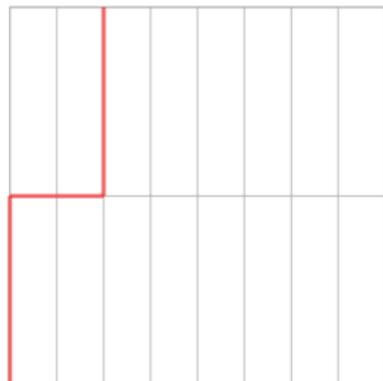
Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



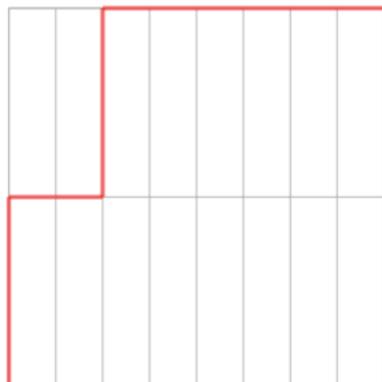
Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



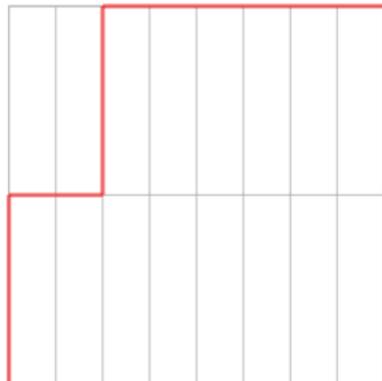
Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



Sortiert nach S_1 bzw. S_2 :

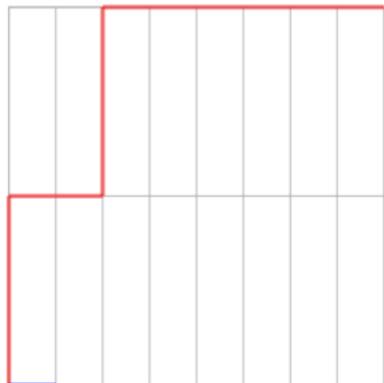
Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



$$AUC(ROC(S_1)) = 14/16 = 0.875$$

Sortiert nach S_1 bzw. S_2 :

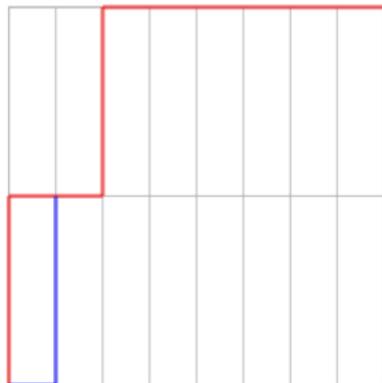
Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



$$AUC(ROC(S_1)) = 14/16 = 0.875$$

Sortiert nach S_1 bzw. S_2 :

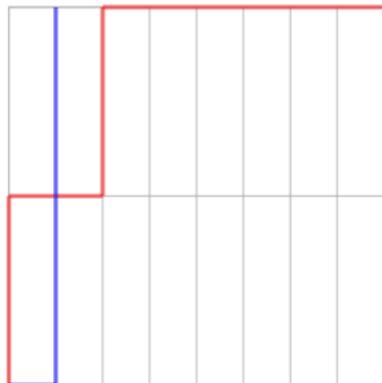
Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



$$AUC(ROC(S_1)) = 14/16 = 0.875$$

Sortiert nach S_1 bzw. S_2 :

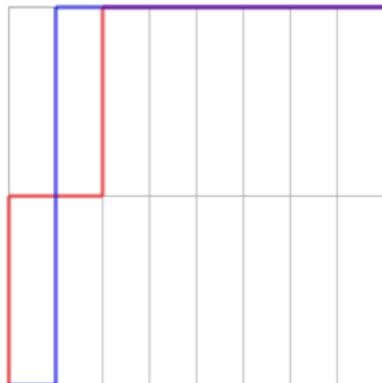
Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



$$AUC(ROC(S_1)) = 14/16 = 0.875$$

Sortiert nach S_1 bzw. S_2 :

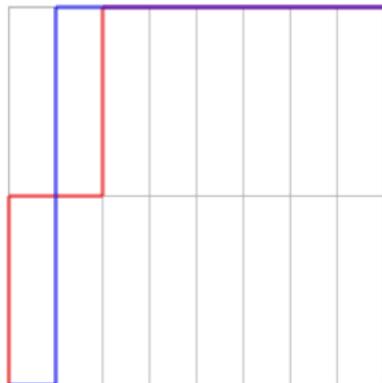
Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



$$AUC(ROC(S_1)) = 14/16 = 0.875$$

Sortiert nach S_1 bzw. S_2 :

Label	+	-	-	+	-	-	-	-	-	-
S_1	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.1	1.0	0.9
Label	-	+	+	-	-	-	-	-	-	-
S_2	.91	.90	.89	.81	.80	.80	.50	.50	.20	.10



$$AUC(ROC(S_1)) = 14/16 = 0.875$$

$$AUC(ROC(S_2)) = 14/16 = 0.875$$

Idee: Scores sollten nicht nur als Rangfolge mit einfließen.

Aufgabe 11-1

Aufgabe 11-2

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Idee: Scores sollten nicht nur als Rangfolge mit einfließen.

Wenn die Scores zwischen 0 (inlier) und 1 (outlier) liegen, kann man den "Fehler" messen.

Aufgabe 11-1

Aufgabe 11-2

C1 und L1

C2 und L2

C3 und L3

C4 und L4

Assoziationsregeln

Idee: Scores sollten nicht nur als Rangfolge mit einfließen.

Wenn die Scores zwischen 0 (inlier) und 1 (outlier) liegen, kann man den "Fehler" messen.

Gewichtung notwendig, da $|O| \ll |I|$.

Heuristik: $\frac{1}{2}$ Gewicht auf Inlier verteilen,
 $\frac{1}{2}$ Gewicht auf die Outlier.

Publiziert in: :-)

H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek.

"Interpreting and Unifying Outlier Scores". In:
*Proceedings of the 11th SIAM International Conference on
Data Mining (SDM), Mesa, AZ. 2011, pp. 13–24*

Lineare Skalierung auf $[0; 1]$:

Label	-	-	-	-	+	-	-	-	+	-
$n(S_1)$.0476	.0952	.0952	.1904	1	.5238	.2857	0	.2381	.1429
$n(S_2)$.8642	.8642	0	.8765	.9753	.4938	.4938	1	.9877	.1235

Lineare Skalierung auf $[0; 1]$:

Label	-	-	-	-	+	-	-	-	+	-
$n(S_1)$.0476	.0952	.0952	.1904	1	.5238	.2857	0	.2381	.1429
$n(S_2)$.8642	.8642	0	.8765	.9753	.4938	.4938	1	.9877	.1235

Kosten Inlier: $\frac{1}{|I|} \sum_{p \in I} n(S_1(p))$ Kosten Outlier: $\frac{1}{|O|} \sum_{p \in O} 1 - n(S_1(p))$

Lineare Skalierung auf $[0; 1]$:

Label	-	-	-	-	+	-	-	-	+	-
$n(S_1)$.0476	.0952	.0952	.1904	1	.5238	.2857	0	.2381	.1429
$n(S_2)$.8642	.8642	0	.8765	.9753	.4938	.4938	1	.9877	.1235

Kosten Inlier: $\frac{1}{|I|} \sum_{p \in I} n(S_1(p))$ Kosten Outlier: $\frac{1}{|O|} \sum_{p \in O} 1 - n(S_1(p))$

	Kosten Inlier	Kosten Outlier	Mittelwert
S_1	.1726	.3810	.2768
S_2	.5895	.0185	.3040

Lineare Skalierung auf $[0; 1]$:

Label	-	-	-	-	+	-	-	-	+	-
$n(S_1)$.0476	.0952	.0952	.1904	1	.5238	.2857	0	.2381	.1429
$n(S_2)$.8642	.8642	0	.8765	.9753	.4938	.4938	1	.9877	.1235

Kosten Inlier: $\frac{1}{|I|} \sum_{p \in I} n(S_1(p))$

Kosten Outlier: $\frac{1}{|O|} \sum_{p \in O} 1 - n(S_1(p))$

	Kosten Inlier	Kosten Outlier	Mittelwert
S_1	.1726	.3810	.2768
S_2	.5895	.0185	.3040

S_2 deutlich geringere Kosten auf den Ausreißern, aber dafür auch mehr "false positives".

Wiederholung:

- ▶ Frequent itemsets sind *typische* Teilmengen
- ▶ Assoziationsregeln *erweitern* ein kleines FI in ein größeres FI
- ▶ Apriori ist der bekannteste Algorithmus hierfür, und ein Verfahren, das *kennen muss* (im data mining)
- ▶ Das Prinzip wird auch außerhalb der Warenkorbanalyse verwendet!
(Das ist nur das Paradebeispiel)
- ▶ Mehr auf den Vorlesungsfolien!

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	
B	
C	
D	
E	
F	
G	
H	
I	
J	
K	
L	

Database

→ 1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	
B	1
C	
D	
E	1
F	
G	1
H	1
I	
J	
K	
L	

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	1
B	2
C	1
D	
E	2
F	
G	2
H	2
I	
J	
K	
L	

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	2
B	3
C	2
D	
E	3
F	1
G	2
H	3
I	
J	
K	
L	

Database

1	B E G H
2	A B C E G H
3	A B C E F H
→ 4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	2
B	4
C	3
D	1
E	4
F	2
G	3
H	4
I	
J	
K	
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
→ 5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	3
B	5
C	3
D	1
E	5
F	2
G	3
H	5
I	
J	
K	1
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
→ 6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	3
B	6
C	3
D	1
E	6
F	3
G	4
H	6
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	4
B	7
C	3
D	2
E	6
F	3
G	5
H	7
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	5
B	8
C	3
D	3
E	6
F	3
G	6
H	7
I	1
J	
K	2
L	1



Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	5
B	9
C	3
D	4
E	6
F	4
G	7
H	7
I	1
J	
K	2
L	1



Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C1

A	5
B	9
C	4
D	4
E	7
F	5
G	7
H	7
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C1

A	6
B	9
C	5
D	4
E	8
F	6
G	7
H	8
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C1

A	7
B	10
C	5
D	4
E	9
F	6
G	8
H	8
I	1
J	
K	2
L	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A	7
B	10
C	5
D	4
E	9
F	6
G	8
H	8
I	4
J	0
K	2
L	4

Minsupport 30% \Leftrightarrow Support $\geq 4/12$

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB		CE	
AC		CF	
AD		CG	
AE		CH	
AF		DE	
AG		DF	
AH		DG	
BC		DH	
BD		EF	
BE		EG	
BF		EH	
BG		FG	
BH		FH	
CD		GH	

Database

→ 1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB		CE	
AC		CF	
AD		CG	
AE		CH	
AF		DE	
AG		DF	
AH		DG	
BC		DH	
BD		EF	
BE	1	EG	1
BF		EH	1
BG	1	FG	
BH	1	FH	
CD		GH	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB	1	CE	1
AC	1	CF	
AD		CG	1
AE	1	CH	1
AF		DE	
AG	1	DF	
AH	1	DG	
BC	1	DH	
BD		EF	
BE	2	EG	2
BF		EH	2
BG	2	FG	
BH	2	FH	
CD		GH	2

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	2	CE	2
AC	2	CF	1
AD		CG	1
AE	2	CH	2
AF	1	DE	
AG	1	DF	
AH	2	DG	
BC	2	DH	
BD		EF	1
BE	3	EG	2
BF	1	EH	3
BG	2	FG	
BH	3	FH	1
CD		GH	2

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	2	CE	3
AC	2	CF	2
AD		CG	2
AE	2	CH	3
AF	1	DE	1
AG	1	DF	1
AH	2	DG	1
BC	3	DH	1
BD	1	EF	2
BE	4	EG	3
BF	2	EH	4
BG	3	FG	1
BH	4	FH	2
CD	1	GH	3

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	3	CE	3
AC	2	CF	2
AD		CG	2
AE	3	CH	3
AF	1	DE	1
AG	1	DF	1
AH	3	DG	1
BC	3	DH	1
BD	1	EF	2
BE	5	EG	3
BF	2	EH	5
BG	3	FG	1
BH	5	FH	2
CD	1	GH	3

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	3	CE	3
AC	2	CF	2
AD		CG	2
AE	3	CH	3
AF	1	DE	1
AG	1	DF	1
AH	3	DG	1
BC	3	DH	1
BD	1	EF	3
BE	6	EG	4
BF	3	EH	6
BG	4	FG	2
BH	6	FH	3
CD	1	GH	4

Database

1	B E G H	
2	A B C E G H	
3	A B C E F H	
4	B C D E F G H L	
5	A B E K H	
6	B E F G H I K	
→	7	A B D G H
8	A B D G	
9	B D F G	
10	C E F	
11	A C E F H	
12	A B E G	

L1

A
B
C
D
E
F
G
H

C2

AB	4	CE	3
AC	2	CF	2
AD	1	CG	2
AE	3	CH	3
AF	1	DE	1
AG	2	DF	1
AH	4	DG	2
BC	3	DH	2
BD	2	EF	3
BE	6	EG	4
BF	3	EH	6
BG	5	FG	2
BH	7	FH	3
CD	1	GH	5

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	5	CE	3
AC	2	CF	2
AD	2	CG	2
AE	3	CH	3
AF	1	DE	1
AG	3	DF	1
AH	4	DG	3
BC	3	DH	2
BD	3	EF	3
BE	6	EG	4
BF	3	EH	6
BG	6	FG	2
BH	7	FH	3
CD	1	GH	5

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB	5	CE	3
AC	2	CF	2
AD	2	CG	2
AE	3	CH	3
AF	1	DE	1
AG	3	DF	2
AH	4	DG	4
BC	3	DH	2
BD	4	EF	3
BE	6	EG	4
BF	4	EH	6
BG	7	FG	3
BH	7	FH	3
CD	1	GH	5

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
→ 10	C E F
11	A C E F H
12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB	5	CE	4
AC	2	CF	3
AD	2	CG	2
AE	3	CH	3
AF	1	DE	1
AG	3	DF	2
AH	4	DG	4
BC	3	DH	2
BD	4	EF	4
BE	6	EG	4
BF	4	EH	6
BG	7	FG	3
BH	7	FH	3
CD	1	GH	5

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



L1

A
B
C
D
E
F
G
H

C2

AB	5	CE	5
AC	3	CF	4
AD	2	CG	2
AE	4	CH	4
AF	2	DE	1
AG	3	DF	2
AH	5	DG	4
BC	3	DH	2
BD	4	EF	5
BE	6	EG	4
BF	4	EH	7
BG	7	FG	3
BH	7	FH	4
CD	1	GH	5

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
→ 12	A B E G

L1

A
B
C
D
E
F
G
H

C2

AB	6	CE	5
AC	3	CF	4
AD	2	CG	2
AE	5	CH	4
AF	2	DE	1
AG	4	DF	2
AH	5	DG	4
BC	3	DH	2
BD	4	EF	5
BE	7	EG	5
BF	4	EH	7
BG	8	FG	3
BH	7	FH	4
CD	1	GH	5

Database

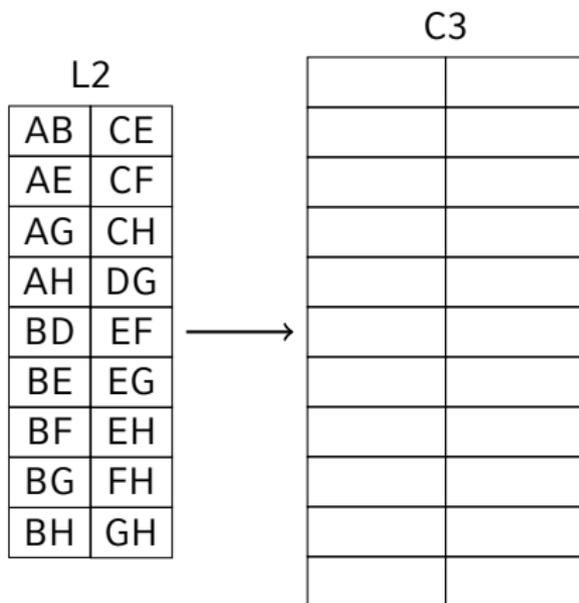
1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L1

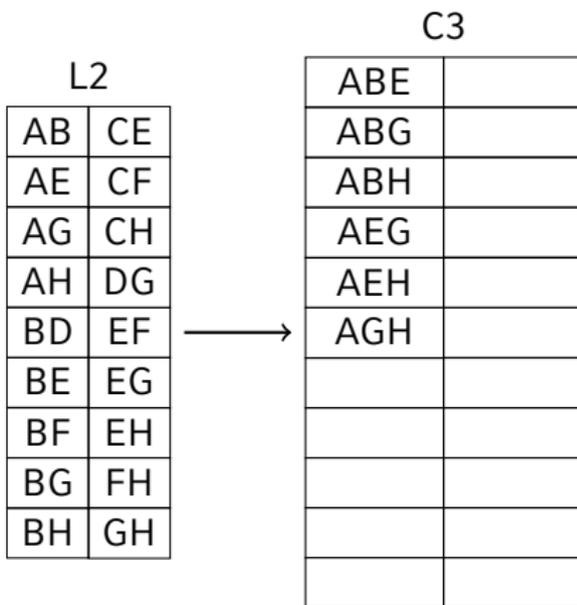
A
B
C
D
E
F
G
H

L2

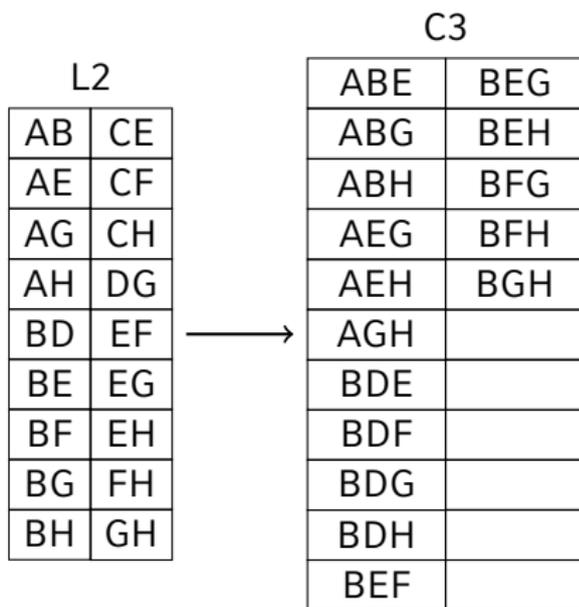
AB	6	CE	5
AC	3	CF	4
AD	2	CG	2
AE	5	CH	4
AF	2	DE	1
AG	4	DF	2
AH	5	DG	4
BC	3	DH	2
BD	4	EF	5
BE	7	EG	5
BF	4	EH	7
BG	8	FG	3
BH	7	FH	4
CD	1	GH	5



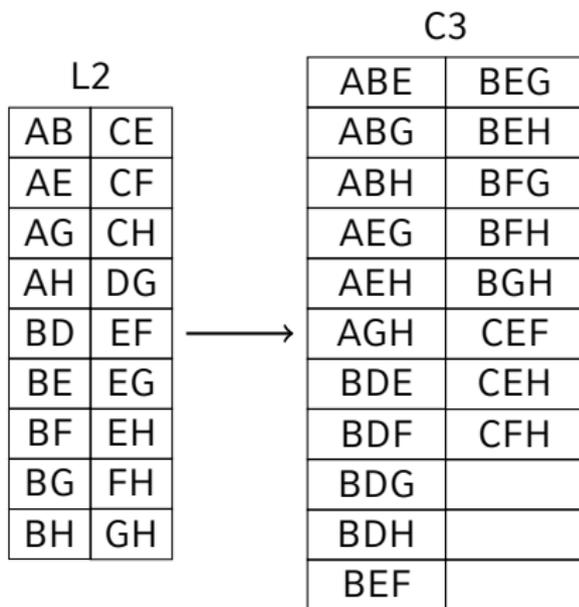
Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!



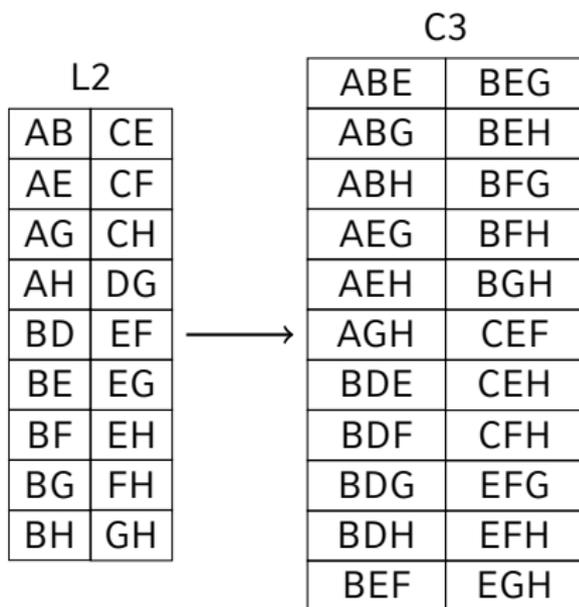
Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!



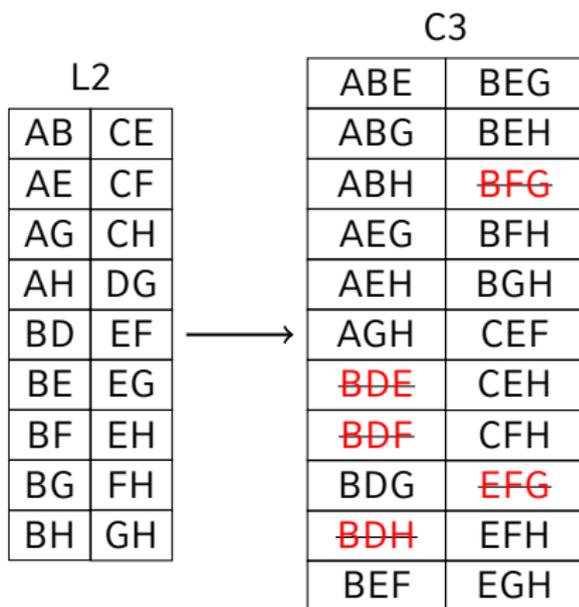
Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!



Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!



Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!



Diese Konstruktion ist der wichtigste Schritt in Apriori!
Ihr werdet dies in vielen "Apriori-style" Algorithmen
wiederfinden!

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE		BEH	
ABG		BFH	
ABH		BGH	
AEG		CEF	
AEH		CEH	
AGH		CFH	
BDG		EFH	
BEF		EGH	
BEG			

Database

→ 1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE		BEH	1
ABG		BFH	
ABH		BGH	1
AEG		CEF	
AEH		CEH	
AGH		CFH	
BDG		EFH	
BEF		EGH	1
BEG	1		

Database

1	B E G H
→ 2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE	1	BEH	2
ABG	1	BFH	
ABH	1	BGH	2
AEG	1	CEF	
AEH	1	CEH	1
AGH	1	CFH	
BDG		EFH	
BEF		EGH	2
BEG	2		

Database

1	B E G H
2	A B C E G H
→ 3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE	2	BEH	3
ABG	1	BFH	1
ABH	2	BGH	2
AEG	1	CEF	1
AEH	2	CEH	2
AGH	1	CFH	1
BDG		EFH	1
BEF	1	EGH	2
BEG	2		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
→ 4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE	2	BEH	4
ABG	1	BFH	2
ABH	2	BGH	3
AEG	1	CEF	2
AEH	2	CEH	3
AGH	1	CFH	2
BDG	1	EFH	2
BEF	2	EGH	3
BEG	3		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
→ 5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C3

ABE	3	BEH	5
ABG	1	BFH	2
ABH	3	BGH	3
AEG	1	CEF	2
AEH	3	CEH	3
AGH	1	CFH	2
BDG	1	EFH	2
BEF	2	EGH	3
BEG	3		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C3

ABE	3	BEH	6
ABG	1	BFH	3
ABH	3	BGH	4
AEG	1	CEF	2
AEH	3	CEH	3
AGH	1	CFH	2
BDG	1	EFH	3
BEF	3	EGH	4
BEG	4		

Database

1	B E G H	
2	A B C E G H	
3	A B C E F H	
4	B C D E F G H L	
5	A B E K H	
6	B E F G H I K	
→	7	A B D G H
8	A B D G	
9	B D F G	
10	C E F	
11	A C E F H	
12	A B E G	

C3

ABE	3	BEH	6
ABG	2	BFH	3
ABH	4	BGH	5
AEG	1	CEF	2
AEH	3	CEH	3
AGH	2	CFH	2
BDG	2	EFH	3
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C3

ABE	3	BEH	6
ABG	3	BFH	3
ABH	4	BGH	5
AEG	1	CEF	2
AEH	3	CEH	3
AGH	2	CFH	2
BDG	3	EFH	3
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C3

ABE	3	BEH	6
ABG	3	BFH	3
ABH	4	BGH	5
AEG	1	CEF	2
AEH	3	CEH	3
AGH	2	CFH	2
BDG	4	EFH	3
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
→ 10	C E F
11	A C E F H
12	A B E G

C3

ABE	3	BEH	6
ABG	3	BFH	3
ABH	4	BGH	5
AEG	1	CEF	3
AEH	3	CEH	3
AGH	2	CFH	2
BDG	4	EFH	3
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C3

ABE	3	BEH	6
ABG	3	BFH	3
ABH	4	BGH	5
AEG	1	CEF	4
AEH	4	CEH	4
AGH	2	CFH	3
BDG	4	EFH	4
BEF	3	EGH	4
BEG	4		

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C3

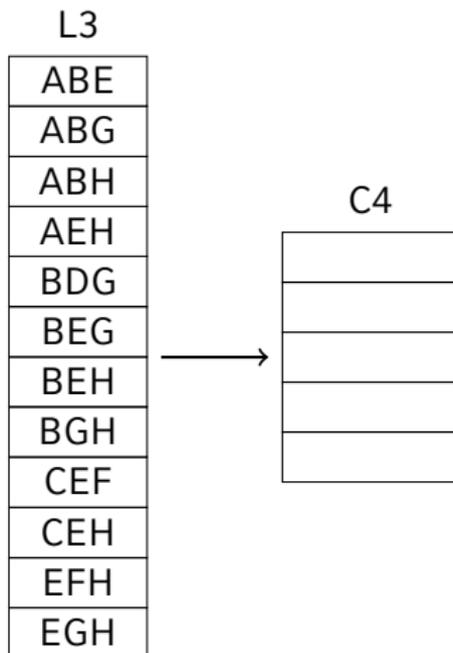
ABE	4	BEH	6
ABG	4	BFH	3
ABH	4	BGH	5
AEG	2	CEF	4
AEH	4	CEH	4
AGH	2	CFH	3
BDG	4	EFH	4
BEF	3	EGH	4
BEG	5		

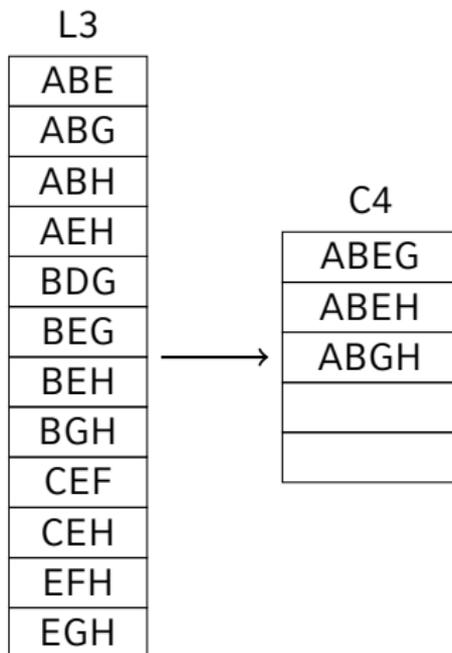
Database

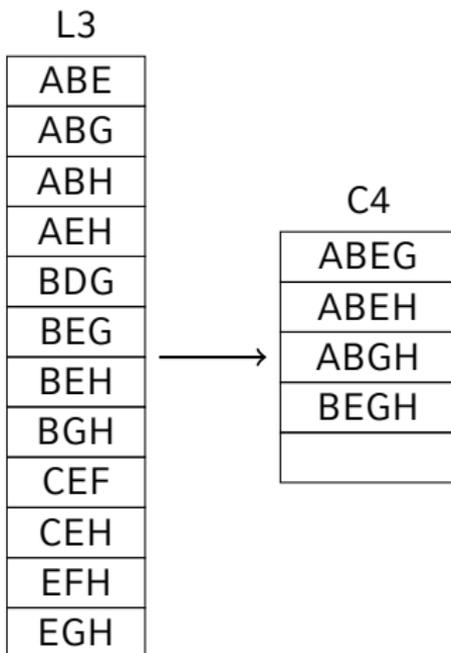
1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

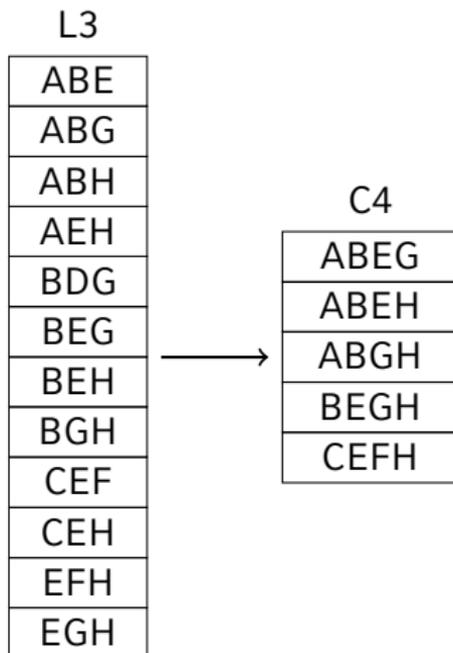
L3

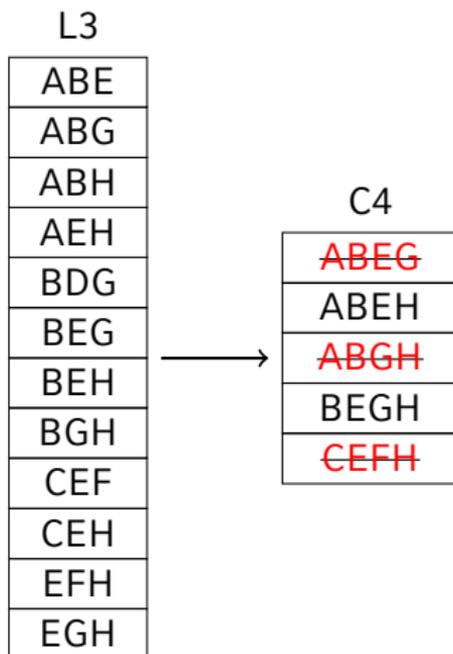
ABE	4	BEH	6
ABG	4	BFH	3
ABH	4	BGH	5
AEG	2	CEF	4
AEH	4	CEH	4
AGH	2	CFH	3
BDG	4	EFH	4
BEF	3	EGH	4
BEG	5		











AEG, AGH und CFH sind nicht *frequent*!

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	
BEGH	

Database

→ 1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	
BEGH	1

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C4

ABEH	1
BEGH	2

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G



C4

ABEH	2
BEGH	2

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	2
BEGH	3

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
→ 5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	3
BEGH	3

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

C4

ABEH	3
BEGH	4

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
→ 7	A B D G H
→ 8	A B D G
→ 9	B D F G
→ 10	C E F
→ 11	A C E F H
→ 12	A B E G

C4

ABEH	3
BEGH	4

Database

1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

L4

ABEH	3
BEGH	4

Nur ein frequent 4-Itemset übrig!

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G
BH	EG	7	$4/7 \approx 0.571$	-

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G
BH	EG	7	$4/7 \approx 0.571$	-
EH	BG	7	$4/7 \approx 0.571$	-

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G
BH	EG	7	$4/7 \approx 0.571$	-
EH	BG	7	$4/7 \approx 0.571$	-
BGH	E	5	$4/5 = \mathbf{0.800}$	BGH \Rightarrow E

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G
BH	EG	7	$4/7 \approx 0.571$	-
EH	BG	7	$4/7 \approx 0.571$	-
BGH	E	5	$4/5 = \mathbf{0.800}$	BGH \Rightarrow E
GH	BE	5	$4/5 = \mathbf{0.800}$	GH \Rightarrow BE

Assoziationsregeln für BEGH, confidence $\geq 60\%$, support 4:

Body	Head	Support	Confidence	Rule
BEGH	\emptyset	4	1.000	
BEG	H	5	$4/5 = \mathbf{0.800}$	BEG \Rightarrow H
BE	GH	7	$4/7 \approx 0.571$	-
BG	EH	8	$4/8 = 0.500$	-
EG	BH	5	$4/5 = \mathbf{0.800}$	EG \Rightarrow BH
BEH	G	6	$4/6 \approx \mathbf{0.667}$	BEH \Rightarrow G
BH	EG	7	$4/7 \approx 0.571$	-
EH	BG	7	$4/7 \approx 0.571$	-
BGH	E	5	$4/5 = \mathbf{0.800}$	BGH \Rightarrow E
GH	BE	5	$4/5 = \mathbf{0.800}$	GH \Rightarrow BE
EGH	B	4	$4/4 = \mathbf{1.000}$	EGH \Rightarrow B