

8. Übung „Knowledge Discovery“

Wintersemester 2008/2009

1 Hierarchische Assoziationsregeln

- Seien A, A_1, A_2, B Items und A eine Verallgemeinerung von A_1 und A_2 . Beweisen oder widerlegen Sie (durch Angabe eines Gegenbeispiels) die folgenden Behauptungen:
 - $\text{support}(A \rightarrow B) = \text{support}(A_1 \rightarrow B) + \text{support}(A_2 \rightarrow B)$
 - Falls $\text{support}(A_1 \rightarrow B) > \text{minsupport}$, dann ist $\text{support}(A \rightarrow B) > \text{minsupport}$.
 - Falls $\text{support}(A \rightarrow B) > \text{minsupport}$, dann ist $\text{support}(A_1 \rightarrow B) > \text{minsupport}$.
- Wie berechnet sich die erwartete Konfidenz einer Regel $X \rightarrow Y$?
- Gegeben folgende Menge von Transaktionen:

TID	Items
1	D,F,G
2	C,D,E,G
3	A,D,E
4	C,D,H
5	B,C,D,F,G
6	C,F,G
7	C,D,F
8	E,G

- Geben Sie für einen Support von 25% die vom Apriori-Algorithmus generierten Mengen C_k und L_k an.
- Bestimmen Sie Support und Konfidenz der Assoziationsregel $DF \rightarrow G$.
- Nehmen Sie an, dass die Items E, F und G zum Item X generalisiert sind. Bestimmen Sie den Support von $X \rightarrow C$. Geben Sie an, ob $X \rightarrow C$ und $E \rightarrow C$ R -interessant mit $R = 2$ sind. Begründen Sie ihre Antworten.

2 TITANIC

Folgend ist der TITANIC-Algorithmus in einer Form wiedergegeben, die der Form des Apriori-Algorithmus aus der Vorlesung ähnelt.

```

1 TITANIC ( $I, D, minsup$ )
2   support( $\{\emptyset\}$ );
3    $L_0 := \{\emptyset\}$ ;
4    $k := 1$ ;
5   forall  $m \in M$  do  $\{m\}.p\_s := \emptyset.s$ ;
6    $C := \{\{m\} \mid m \in M\}$ ; // Singletons
7   loop begin
8     support( $C$ ); // Support berechnen
9     forall  $X \in L_{k-1}$  do  $X.closure := closure(X)$ ;
10     $L_k := \{X \in C \mid X.s \neq X.p\_s \wedge X.s \geq minsup\}$ ; // Pruning
11    if  $L_k = \emptyset$  then exit loop;
12     $k ++$ ;
13     $C :=$  Apriori-Kandidatengenerierung ( $L_{k-1}$ );
14  end loop;
15  return  $\bigcup_{i=0}^{k-1} \{X.closure \mid X \in L_i\}$ ;

```

1. Berechnen Sie für den Kontext auf Seite 12 der Vorlesungs-Folien (Kapitel 4, Teil 2 Begriffsverbände) die Begriffsinhalte mit einem Mindest-Support von 1 (also $33\frac{1}{3}\%$).
2. Stellen Sie den Zusammenhang zwischen Ihren Zwischenergebnissen und den auf der Folie 12 genannten Begriffen und Formeln her.
3. Vergleichen Sie die einzelnen Schritte (Pruning, Kandidatengenerierung, Hüllen) des TITANIC- mit denen des Apriori-Algorithmus.
4. Diskutieren Sie für jeden der angegebenen Teilschritte die möglichen Performance-Gewinne und -Verluste.