

Beschreibung der fachlichen Lösung

- Wir bezeichnen die Menge aller Transaktionen ("Einkäufe") als D .
Eine Transaktion ist dabei eine nicht-leere Menge von Artikeln. Damit gilt:
 $D \subseteq (P(\text{Artikel}) \setminus \{\})$, wobei D eine Multimenge ist. Weiterhin gilt:
 $\forall t \in D : t \in (P(\text{Artikel}) \setminus \{\})$.

- Zum Beispiel:

$$\begin{aligned}\text{Artikel} &= \{A1, A2, A3\} \\ P(\text{Artikel}) &= \{\{\}, \{A1\}, \{A2\}, \{A3\}, \{A1, A2\}, \\ &\quad \{A1, A3\}, \{A2, A3\}, \{A1, A2, A3\}\} \\ D &= \{\{A1\}, \{A1, A2\}, \{A1, A2\}, \{A2, A3\}\}\end{aligned}$$

- Beispielsweise berechnet sich für zwei beliebige Artikel x und y der Parameter Konfidenz der Assoziationsregel $\{x\} \rightarrow \{y\}$ wie folgt:

$$\frac{|\{t \in D : \{x\} \cup \{y\} \subseteq t\}|}{|\{u \in D : \{x\} \subseteq u\}|} = \frac{|\{t \in D : \{x, y\} \subseteq t\}|}{|\{u \in D : \{x\} \subseteq u\}|}$$

In Worten:

$$\frac{\text{Anzahl aller Transaktionen die } x \text{ und } y \text{ enthalten}}{\text{Anzahl aller Transaktionen die } x \text{ enthalten}}$$

- Für zwei beliebige Artikel x und y berechnet sich der Parameter Support der Assoziationsregel $\{x\} \rightarrow \{y\}$ wie folgt:

$$\frac{|\{t \in D : \{x\} \cup \{y\} \subseteq t\}|}{|D|} = \frac{|\{t \in D : \{x, y\} \subseteq t\}|}{|D|}$$

In Worten:

$$\frac{\text{Anzahl aller Transaktionen die } x \text{ und } y \text{ enthalten}}{\text{Anzahl aller Transaktionen}}$$

- Diese Vorgehensweise lässt sich auch bei beliebigen Teilmengen der Menge Artikel anwenden.
- Um alle Teilmengen zu bestimmen, die den minimalen Support erfüllen, bilden wir zuerst alle einelementrigen Teilmengen, die den minimalen Support erfüllen. Diese Menge bezeichnen wir nun als `articles`.
Danach speichern wir die Menge `articles` in der Menge `results`.
Im folgenden iterieren wir über `results`, solange bis keine Elemente mehr dazu kommen, wobei auch die neu hinzugekommenen Elemente dabei miteinbezogen werden. Bei der Iteration über `results` wird für jedes Element von `results` die Vereinigung mit einem Element aus `articles` gebildet und diese neue Menge zu der Menge `results` hinzugefügt, solange der minimale Support auch für diese Menge erfüllt wird.
- Wir bestimmen nun für Elemente der Menge `results` und für alle Partitionen dieser Elemente die Ergebnisse der Assoziationsanalyse, d.h. wir berechnen Support und Konfidenz. Jedoch fügen wir jede Partionen der Menge nur dann das Ergebnis zum Endergebnis hinzu, wenn auch die minimale Konfidenz erfüllt wurde. Den minimalen Support müssen wir nicht mehr prüfen, da dieser für alle Partionen des Elements gleich ist.

Beschreibung der Implementierung

- **Funktion 'get_ssociation_analysis_data':**

Um die Assoziationsanalyse zu bestimmen, definieren wir uns eine Funktion `association_analysis_data(min_support, min_confidence)`.

Die Funktion liefert für alle Assoziationsregeln als Rückgabewert folgende Werte:

1. Artikel aus x (als Array, auch `articles_antecedent`)
2. Artikel aus y (als Array, auch `articles_consequent`)
3. Anzahl der Transaktionen die x enthalten
4. Anzahl der Transaktionen die y enthalten
5. Anzahl der Transaktionen die x und y enthalten
6. Parameter Konfidenz
7. Parameter Support.

Die einzelnen Assoziationsregeln sind dabei die Elemente eines Arrays, das zurückgegeben wird.

Zuerst werden nach dem Apriori-Algorithmus alle Sets von Artikeln bestimmt, die den minimalen Support erfüllen.

Die Berechnung der Werte ist wie folgt implementiert, wobei auf eine Hilfsfunktion `get_bill_count` zurückgegriffen wird, die im nächsten Abschnitt beschrieben wird. Der Begriff Transaktion ist dabei bei unserem Datenmodell synonym zur Rechnung:

1. Bestimme **Anzahl aller Transaktionen**:
`count_all = Bill.count`
2. Bestimme **Anzahl aller Transaktionen die x enthalten**:
`count_x = get_bill_count(articles_antecedent)`
3. Bestimme **Anzahl aller Transaktionen die y enthalten**:
`count_y = get_bill_count(articles_consequent)`
4. Bestimme **Anzahl aller Transaktionen die x und y enthalten**:
`count_xy = get_bill_count(articles_antecedent | articles_consequent)`
5. Bestimme Konfidenz:
`confidence = count_xy / count_x`
6. Bestimme Support:
`support = count_xy / count_all`
7. Prüfe, ob bei der Berechnung der Parameter Konfidenz und Support eine Division durch 0 auftritt und liefere in diesem Fall als Ergebnis den Wert 0.0.
8. Liefere die Ergebnisse zurück:
`return [count_x, count_y, count_xy, confidence, support]`.

- **Funktion 'get_bill_count':**

Hilfsfunktion mit einem Parameter `articles`, welcher eine Menge von Artikeln darstellt. Die Funktion berechnet die Anzahl aller Transaktionen, die alle Artikel aus `articles` enthalten.

Die einzelnen Transaktionen stellen wir aus Übersichtlichkeitsgründen nicht weiter dar. Wir haben jedoch zur Erweiterbarkeit eine Funktion `get_bills(articles)` definiert, die uns die entsprechenden Transaktionen (Rechnungen) als Array zurückliefert. `get_bill_count(articles)` greift dabei auf diese Funktion zurück und liefert die Anzahl des zurückgelieferten Arrays zurück:

```
return get_bills(articles).count
```

Die Funktion `get_bills(articles)` ist wie folgt implementiert:

1. Iteriere über jedes Element aus `articles` und speichere für jedes Element alle

Rechnungen die das jeweilige Element enthalten in Array. Speichere diese Arrays nun in ein übergeordnetes Array:

```
article_bills = articles.map{|article|
```

```
BillEntry.where(:article => article).map{|entry|  
  entry.bill}.uniq}
```

2. Iteriere über das übergeordnete Array und bilde dabei von jedem Element den Durchschnitt:

```
result = article_bills.reduce([])  
  {|accu, entry| accu & entry}
```

3. Liefere das Ergebnis zurück:

```
return result
```

- Wir stellen die numerische Ergebnisse unserer Analyse auf der Unterseite /associations dar. In der Anzeige bestimmen unter Berücksichtigung der Parameter 'minimaler Support' und 'minimale Konfidenz' die Analyseergebnisse aus. Dabei werden Artikelnummern der linken und rechten Seite der Assoziationsregel und der Werte für Support und Konfidenz ausgegeben.

Die Ergebnisse werden zuerst nach Konfidenz und danach nach Support sortiert.

Weiterhin wird in absoluten Zahlen angegeben, wieviele Transaktionen insgesamt getätigt wurden.

- Besucht ein Gast oder ein angemeldeter Benutzer, der kein Administrator ist, die Seite, wird kein Inhalt auf der Seite angezeigt.