

Knowledge Discovery in Databases
SS 2013

Übungsblatt 11: Evaluation von Outlier-Detection-Algorithmen, Apriori

Aufgabe 11-1 Evaluation von Outlier-Detection-Algorithmen

Auf einen Datensatz mit bekannten Ausreißern + wurden zwei Verfahren S_1 und S_2 angewendet. Die Ergebnisse der Verfahren finden Sie in folgender Tabelle:

Object	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
Label	-	-	-	-	+	-	-	-	+	-
S_1	1.0	1.1	1.1	1.3	3.0	2.0	1.5	0.9	1.4	1.2
S_2	.80	.80	.10	.81	.89	.50	.50	.91	.90	.20

Bewerten Sie die beiden Ausreißerverfahren S_1 und S_2 mit den folgenden Metriken:

- Precision, Recall und F-Measure, unter der Annahme, dass die größten $k = 2$ Werte als Ausreißer klassifiziert werden.
- Average Precision für $k = [1 \dots 4]$, unter der Annahme, dass die größten k Werte als Ausreißer klassifiziert werden.
- Zeichnen Sie die ROC Kurve, und berechnen sie die Fläche unter dieser Kurve (AUC).
- Normalisieren Sie die Werte auf $[0; 1]$, und berechnen Sie die Kostenfunktion aus der Vorlesung.

Aufgabe 11-2 Apriori-Algorithmus

Gegeben ist die Menge der Items $I = \{A, B, C, D, E, F, G, H, I, K, L, M\}$.

Weiterhin ist eine Menge von Transaktionen T laut folgender Tabelle gegeben:

Menge der Transaktionen T

Transaktions ID	gekaufte Items
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

- (a) Bestimmen Sie zum minimalen Support von 30% die häufig auftretenden Itemsets. Verwenden Sie dazu den Apriori-Algorithmus. Geben Sie insbesondere die Kandidatenmengen nach den Join-Schritten und nach den Prune-Schritten an, sowie die häufig auftretenden Itemsets mit ihrem jeweiligen Support.
- (b) Bestimmen sie alle Assziationregeln, die man aus dem frequent Itemset $\{B E G H\}$ mit einer Konfidenz von mindestens 60% und dem Support 4 (Frequency 30%) bilden kann. Nutzen sie dabei die in der Vorlesung vorgestellte Monotonie aus.