Algorithms and Applications of Data Mining

Yijun Lin
yijunlin@usc.edu

02/13

Finding Frequent Itemset

Consider the following input file of basket data and a support threshold **s = 2**, answer the following questions.

$B_1 = \{m, c, b\}$	$B_2 = \{m, p, j\}$
$B_3 = \{m, c, b, n\}$	$B_4 = \{c, j\}$
$B_5 = \{m, p, b\}$	$B_6 = \{m, c, b, j\}$
$B_7 = \{c, b, j\}$	$B_8 = \{b, c\}$

Find all frequent itemsets with set size <= 3

Write down one association rule and its confidence and interest numbers. Your association rule should be derived from a frequent pair

Example: {m, c} -> {b}

Confidence: Ratio of support for I U {j} with support for I

Interest: $\text{Interest}(I \rightarrow j) = \text{conf}(I \rightarrow j) - \Pr[j]$

Finding Frequent Itemset

Consider the following input file of basket data and a support threshold **s = 2**, answer the following questions.

$B_1 = \{m, c, b\}$ $B_2 = \{m, p, j\}$

$B_3 = \{m, c, b, n\}$ $B_4 = \{c, j\}$

$B_5 = \{m, p, b\}$ $B_6 = \{m, c, b, j\}$

$B_7 = \{c, b, j\}$ $B_8 = \{b, c\}$

Find all frequent itemsets with set size <= 3

{m}, {c}, {b}, {p}, {j}
{m, c}, {m, b}, {m, p}, {b, j}, {m, j}, {c, b}, {c, j}
{m, c, b}, {c, b, j}

Write down one association rule and its confidence and interest numbers. Your association rule should be derived from a frequent pair

Example: {m, c} -> {b}

confidence : 1 interest: 1 - 6/8 = 0.25

Confidence: Ratio of support for I U {j} with support for I

Interest: $\text{Interest}(I \rightarrow j) = \text{conf}(I \rightarrow j) - \Pr[j]$

Finding Frequent Itemset

Here is a collection of 6 baskets. Each contains three of the six items 1 through 6.
{1, 2, 3 } {2, 3, 4 } {3, 4, 5 } {4, 5, 6 }{1, 3, 5 } {2, 4, 6 } The support threshold is 2. The hash function is i x j mod 11. Using the **Apriori Algorithm**, you need to show 1. frequent single items and 2. frequent pairs.

Algorithm 1 Apriori Algorithm

1: *Apriori*(T, σ)
2: $L_1 \leftarrow \{\text{large1} - \text{itemsets}\}$
3: $k \leftarrow 2$
4: **while** $L_k \neq emptyset$
5: $C_k \leftarrow \left\{a \cup \{\mathrm{b}\} \mid \mathrm{a} \in L_{k-1} \wedge b \in \bigcup L_{k-1} \wedge b \notin a\right\}$
6: **for** *transactions* $t \in T$
7: $C_t \leftarrow \left\{c \mid c \in C_k \wedge c \subseteq t\right\}$
8: **for** *candidates* $c \in C_t$
9: $count[c] \leftarrow \text{count}[c] + 1$
10: $L_k \leftarrow \left\{c \mid c \in C_k \wedge count[c] \geq \sigma\right\}$
11: $k \leftarrow k + 1$
12: **return** $\bigcup_k L_k$

Finding Frequent Itemset

Here is a collection of 6 baskets. Each contains three of the six items 1 through 6.
{1, 2, 3 } {2, 3, 4 } {3, 4, 5 } {4, 5, 6 }{1, 3, 5 } {2, 4, 6 } The support threshold is 2. The hash function is i x j mod 11. Using the **Apriori Algorithm**, you need to show 1. frequent single items and 2. frequent pairs.

Frequent single items: {1},{2},{3},{4},{5},{6}
Frequent Pairs: {1,3}, {2,3}, {2,4}, {3,4}, {3,5}, {4,5} {4,6}