CS4445 B Term 2006
Homework 2 Solutions

Contents

2 Apriori Associacion Rule Generation **1**
2.1 Itemset Generation . . . 1
2.1.1 Level 1 Frequent Itemsets . . . 1
2.1.2 Level 2 Frequent Itemsets . . . 2
2.1.3 Level 3 Frequent Itemsets . . . 5
2.1.4 Level 4 Frequent Itemsets . . . 6
2.2 Association Rule Generation . . . 6
2.2.1 Rules from itemset {**buying-price**=*med*, **safety**=*med*} . . . 6
2.2.2 Rules from itemset {**buying-price**=*med*, **recommendation**=*acc*} . . . 6
2.2.3 Rules from itemset {**buying-price**=*low*, **persons**=*2*} . . . 7
2.2.4 Rules from itemset {**buying-price**=*low*, **safety**=*high*} . . . 7
2.2.5 Rules from itemset {**buying-price**=*low*, **recommendation**=*unacc*} . . . 7
2.2.6 Rules from itemset {**maintenance**=*med*, **persons**=*4*} . . . 7
2.2.7 Rules from itemset {**maintenance**=*vhigh*, **safety**=*med*} . . . 7
2.2.8 Rules from itemset {**persons**=*2*, **recommendation**=*unacc*} . . . 8
2.2.9 Rules from itemset {**persons**=*4*, **safety**=*med*} . . . 8
2.2.10 Rules from itemset {**safety**=*med*, **recommendation**=*acc*} . . . 8
2.2.11 Rules from itemset {**buying-price**=*med*, **safety**=*med*, **recommendation**=*acc*} . . . 8
2.2.12 Rules from itemset {**buying-price**=*low*, **persons**=*2*, **recommendation**=*unacc*} . . . 9
2.3 Final Association Rule List . . . 9

2 Apriori Associacion Rule Generation

2.1 Itemset Generation

We begin with the generation of frequent itemsets.

2.1.1 Level 1 Frequent Itemsets

The candidate level1 itemsets are:

- (2 support) {**buying-price**=*high*}
- (4 support) {**buying-price**=*med*}
- (1 support) {**buying-price**=*vhigh*}
- (5 support) {**buying-price**=*low*}
- (1 support) {**maintenance**=*high*}

- (5 support) {**maintenance**=*med*}
- (4 support) {**maintenance**=*vhigh*}
- (2 support) {**maintenance**=*low*}
- (3 support) {**persons**=*2*}
- (6 support) {**persons**=*4*}
- (3 support) {**persons**=*more*}
- (4 support) {**safety**=*high*}
- (6 support) {**safety**=*med*}
- (2 support) {**safety**=*low*}
- (4 support) {**recommendation**=*acc*}
- (6 support) {**recommendation**=*unacc*}
- (2 support) {**recommendation**=*good*}

We remove those with support less than 3 and we are left with:

- {**buying-price**=*med*}
- {**buying-price**=*low*}
- {**maintenance**=*med*}
- {**maintenance**=*vhigh*}
- {**persons**=*2*}
- {**persons**=*4*}
- {**persons**=*more*}
- {**safety**=*high*}
- {**safety**=*med*}
- {**recommendation**=*acc*}
- {**recommendation**=*unacc*}

2.1.2 Level 2 Frequent Itemsets

We generate the next level by considering itemsets made from merging any two level 1 sets that are the equivalent in the first 0 item(s). To save space we also do not consider itemsets with contradictory values for the same attribute.

- {**buying-price**=*med*} ∪ {**maintenance**=*med*}
  = (1 support) {**buying-price**=*med*, **maintenance**=*med*}
- {**buying-price**=*med*} ∪ {**maintenance**=*vhigh*}
  = (2 support) {**buying-price**=*med*, **maintenance**=*vhigh*}
- {**buying-price**=*med*} ∪ {**persons**=*2*}
  = (0 support) {**buying-price**=*med*, **persons**=*2*}

- {**buying-price**=*med*} ∪ {**persons**=*4*}
  = (2 support) {**buying-price**=*med*, **persons**=*4*}
- {**buying-price**=*med*} ∪ {**persons**=*more*}
  = (2 support) {**buying-price**=*med*, **persons**=*more*}
- {**buying-price**=*med*} ∪ {**safety**=*high*}
  = (0 support) {**buying-price**=*med*, **safety**=*high*}
- {**buying-price**=*med*} ∪ {**safety**=*med*}
  = (3 support) {**buying-price**=*med*, **safety**=*med*}
- {**buying-price**=*med*} ∪ {**recommendation**=*acc*}
  = (3 support) {**buying-price**=*med*, **recommendation**=*acc*}
- {**buying-price**=*med*} ∪ {**recommendation**=*unacc*}
  = (1 support) {**buying-price**=*med*, **recommendation**=*unacc*}
- {**buying-price**=*low*} ∪ {**maintenance**=*med*}
  = (2 support) {**buying-price**=*low*, **maintenance**=*med*}
- {**buying-price**=*low*} ∪ {**maintenance**=*vhigh*}
  = (1 support) {**buying-price**=*low*, **maintenance**=*vhigh*}
- {**buying-price**=*low*} ∪ {**persons**=*2*}
  = (3 support) {**buying-price**=*low*, **persons**=*2*}
- {**buying-price**=*low*} ∪ {**persons**=*4*}
  = (1 support) {**buying-price**=*low*, **persons**=*4*}
- {**buying-price**=*low*} ∪ {**persons**=*more*}
  = (1 support) {**buying-price**=*low*, **persons**=*more*}
- {**buying-price**=*low*} ∪ {**safety**=*high*}
  = (3 support) {**buying-price**=*low*, **safety**=*high*}
- {**buying-price**=*low*} ∪ {**safety**=*med*}
  = (2 support) {**buying-price**=*low*, **safety**=*med*}
- {**buying-price**=*low*} ∪ {**recommendation**=*acc*}
  = (1 support) {**buying-price**=*low*, **recommendation**=*acc*}
- {**buying-price**=*low*} ∪ {**recommendation**=*unacc*}
  = (3 support) {**buying-price**=*low*, **recommendation**=*unacc*}
- {**maintenance**=*med*} ∪ {**persons**=*2*}
  = (1 support) {**maintenance**=*med*, **persons**=*2*}
- {**maintenance**=*med*} ∪ {**persons**=*4*}
  = (3 support) {**maintenance**=*med*, **persons**=*4*}
- {**maintenance**=*med*} ∪ {**persons**=*more*}
  = (1 support) {**maintenance**=*med*, **persons**=*more*}
- {**maintenance**=*med*} ∪ {**safety**=*high*}
  = (2 support) {**maintenance**=*med*, **safety**=*high*}
- {**maintenance**=*med*} ∪ {**safety**=*med*}
  = (2 support) {**maintenance**=*med*, **safety**=*med*}

- {**maintenance**=*med*} ∪ {**recommendation**=*acc*}
  = (1 support) {**maintenance**=*med*, **recommendation**=*acc*}
- {**maintenance**=*med*} ∪ {**recommendation**=*unacc*}
  = (2 support) {**maintenance**=*med*, **recommendation**=*unacc*}
- {**maintenance**=*vhigh*} ∪ {**persons**=*2*}
  = (0 support) {**maintenance**=*vhigh*, **persons**=*2*}
- {**maintenance**=*vhigh*} ∪ {**persons**=*4*}
  = (2 support) {**maintenance**=*vhigh*, **persons**=*4*}
- {**maintenance**=*vhigh*} ∪ {**persons**=*more*}
  = (2 support) {**maintenance**=*vhigh*, **persons**=*more*}
- {**maintenance**=*vhigh*} ∪ {**safety**=*high*}
  = (0 support) {**maintenance**=*vhigh*, **safety**=*high*}
- {**maintenance**=*vhigh*} ∪ {**safety**=*med*}
  = (3 support) {**maintenance**=*vhigh*, **safety**=*med*}
- {**maintenance**=*vhigh*} ∪ {**recommendation**=*acc*}
  = (2 support) {**maintenance**=*vhigh*, **recommendation**=*acc*}
- {**maintenance**=*vhigh*} ∪ {**recommendation**=*unacc*}
  = (2 support) {**maintenance**=*vhigh*, **recommendation**=*unacc*}
- {**persons**=*2*} ∪ {**safety**=*high*}
  = (2 support) {**persons**=*2*, **safety**=*high*}
- {**persons**=*2*} ∪ {**safety**=*med*}
  = (1 support) {**persons**=*2*, **safety**=*med*}
- {**persons**=*2*} ∪ {**recommendation**=*acc*}
  = (0 support) {**persons**=*2*, **recommendation**=*acc*}
- {**persons**=*2*} ∪ {**recommendation**=*unacc*}
  = (3 support) {**persons**=*2*, **recommendation**=*unacc*}
- {**persons**=*4*} ∪ {**safety**=*high*}
  = (2 support) {**persons**=*4*, **safety**=*high*}
- {**persons**=*4*} ∪ {**safety**=*med*}
  = (3 support) {**persons**=*4*, **safety**=*med*}
- {**persons**=*4*} ∪ {**recommendation**=*acc*}
  = (2 support) {**persons**=*4*, **recommendation**=*acc*}
- {**persons**=*4*} ∪ {**recommendation**=*unacc*}
  = (2 support) {**persons**=*4*, **recommendation**=*unacc*}
- {**persons**=*more*} ∪ {**safety**=*high*}
  = (0 support) {**persons**=*more*, **safety**=*high*}
- {**persons**=*more*} ∪ {**safety**=*med*}
  = (2 support) {**persons**=*more*, **safety**=*med*}
- {**persons**=*more*} ∪ {**recommendation**=*acc*}
  = (2 support) {**persons**=*more*, **recommendation**=*acc*}

- {**persons**=*more*} ∪ {**recommendation**=*unacc*}
  = (1 support) {**persons**=*more*, **recommendation**=*unacc*}
- {**safety**=*high*} ∪ {**recommendation**=*acc*}
  = (0 support) {**safety**=*high*, **recommendation**=*acc*}
- {**safety**=*high*} ∪ {**recommendation**=*unacc*}
  = (2 support) {**safety**=*high*, **recommendation**=*unacc*}
- {**safety**=*med*} ∪ {**recommendation**=*acc*}
  = (4 support) {**safety**=*med*, **recommendation**=*acc*}
- {**safety**=*med*} ∪ {**recommendation**=*unacc*}
  = (2 support) {**safety**=*med*, **recommendation**=*unacc*}

We filter out those itemsets with support less than 3 and we are left with:

- {**buying-price**=*med*, **safety**=*med*}
- {**buying-price**=*med*, **recommendation**=*acc*}
- {**buying-price**=*low*, **persons**=*2*}
- {**buying-price**=*low*, **safety**=*high*}
- {**buying-price**=*low*, **recommendation**=*unacc*}
- {**maintenance**=*med*, **persons**=*4*}
- {**maintenance**=*vhigh*, **safety**=*med*}
- {**persons**=*2*, **recommendation**=*unacc*}
- {**persons**=*4*, **safety**=*med*}
- {**safety**=*med*, **recommendation**=*acc*}

2.1.3 Level 3 Frequent Itemsets

We generate the next level by considering itemsets made from merging any two level 2 sets that are the equivalent in the first 1 item(s). To save space we also do not consider itemsets with contradictory values for the same attribute.

- {**buying-price**=*med*, **safety**=*med*} ∪ {**buying-price**=*med*, **recommendation**=*acc*}
  = (3 support) {**buying-price**=*med*, **safety**=*med*, **recommendation**=*acc*}
- {**buying-price**=*low*, **persons**=*2*} ∪ {**buying-price**=*low*, **safety**=*high*}
  = (2 support) {**buying-price**=*low*, **persons**=*2*, **safety**=*high*}
- {**buying-price**=*low*, **persons**=*2*} ∪ {**buying-price**=*low*, **recommendation**=*unacc*}
  = (3 support) {**buying-price**=*low*, **persons**=*2*, **recommendation**=*unacc*}
- {**buying-price**=*low*, **safety**=*high*} ∪ {**buying-price**=*low*, **recommendation**=*unacc*}
  = (2 support) {**buying-price**=*low*, **safety**=*high*, **recommendation**=*unacc*}

We filter out those itemsets with support less than 3 and we are left with:

- {**buying-price**=*med*, **safety**=*med*, **recommendation**=*acc*}
- {**buying-price**=*low*, **persons**=*2*, **recommendation**=*unacc*}

2.1.4 Level 4 Frequent Itemsets

We generate the next level by considering itemsets made from merging any two level 3 sets that are the equivalent in the first 2 item(s). To save space we also do not consider itemsets with contradictory values for the same attribute.

- NO CANDIDATES

Since there are no candidates at this leve, we are done with itemset generation.

2.2 Association Rule Generation

Our frequent itemset generation came up with the following itemsets of size at least two (we are ignoring the ones with one item as they cannot produce any association rules):

- {**buying-price**=*med*, **safety**=*med*}
- {**buying-price**=*med*, **recommendation**=*acc*}
- {**buying-price**=*low*, **persons**=*2*}
- {**buying-price**=*low*, **safety**=*high*}
- {**buying-price**=*low*, **recommendation**=*unacc*}
- {**maintenance**=*med*, **persons**=*4*}
- {**maintenance**=*vhigh*, **safety**=*med*}
- {**persons**=*2*, **recommendation**=*unacc*}
- {**persons**=*4*, **safety**=*med*}
- {**safety**=*med*, **recommendation**=*acc*}
- {**buying-price**=*med*, **safety**=*med*, **recommendation**=*acc*}
- {**buying-price**=*low*, **persons**=*2*, **recommendation**=*unacc*}

We consider each of these in turn to generate association rules.

2.2.1 Rules from itemset {**buying-price=*med*, safety=*med***}

We consider all possible splits of this itemset into antecedent and consequent:

- (75.00% ; $\frac{3}{4}$) {**buying-price**=*med*} ⇒ {**safety**=*med*}
- (50.00% ; $\frac{3}{6}$) {**safety**=*med*} ⇒ {**buying-price**=*med*}

None of these have the sufficient confidence (90.00%) so this itemset does not yield any association rules.

2.2.2 Rules from itemset {**buying-price=*med*, recommendation=*acc***}

We consider all possible splits of this itemset into antecedent and consequent:

- (75.00% ; $\frac{3}{4}$) {**buying-price**=*med*} ⇒ {**recommendation**=*acc*}
- (75.00% ; $\frac{3}{4}$) {**recommendation**=*acc*} ⇒ {**buying-price**=*med*}

None of these have the sufficient confidence (90.00%) so this itemset does not yield any association rules.

2.2.3 Rules from itemset {**buying-price=*low*, persons=*2***}

We consider all possible splits of this itemset into antecedent and consequent:

- (60.00% ; $\frac{3}{5}$) {**buying-price**=*low*} ⇒ {**persons**=*2*}
- (100.00% ; $\frac{3}{3}$) {**persons**=*2*} ⇒ {**buying-price**=*low*}

We filter out those rules below our preset confidence requirement (90.00%) and we are left with the following rules that are added to our ruleset:

- (100.00% ; $\frac{3}{3}$) {**persons**=*2*} ⇒ {**buying-price**=*low*}

2.2.4 Rules from itemset {**buying-price=*low*, safety=*high***}

We consider all possible splits of this itemset into antecedent and consequent:

- (60.00% ; $\frac{3}{5}$) {**buying-price**=*low*} ⇒ {**safety**=*high*}
- (75.00% ; $\frac{3}{4}$) {**safety**=*high*} ⇒ {**buying-price**=*low*}

None of these have the sufficient confidence (90.00%) so this itemset does not yield any association rules.

2.2.5 Rules from itemset {**buying-price=*low*, recommendation=*unacc***}

We consider all possible splits of this itemset into antecedent and consequent:

- (60.00% ; $\frac{3}{5}$) {**buying-price**=*low*} ⇒ {**recommendation**=*unacc*}
- (50.00% ; $\frac{3}{6}$) {**recommendation**=*unacc*} ⇒ {**buying-price**=*low*}

None of these have the sufficient confidence (90.00%) so this itemset does not yield any association rules.

2.2.6 Rules from itemset {**maintenance=*med*, persons=*4***}

We consider all possible splits of this itemset into antecedent and consequent:

- (60.00% ; $\frac{3}{5}$) {**maintenance**=*med*} ⇒ {**persons**=*4*}
- (50.00% ; $\frac{3}{6}$) {**persons**=*4*} ⇒ {**maintenance**=*med*}

None of these have the sufficient confidence (90.00%) so this itemset does not yield any association rules.

2.2.7 Rules from itemset {**maintenance=*vhigh*, safety=*med***}

We consider all possible splits of this itemset into antecedent and consequent:

- (75.00% ; $\frac{3}{4}$) {**maintenance**=*vhigh*} ⇒ {**safety**=*med*}
- (50.00% ; $\frac{3}{6}$) {**safety**=*med*} ⇒ {**maintenance**=*vhigh*}

None of these have the sufficient confidence (90.00%) so this itemset does not yield any association rules.

2.2.8 Rules from itemset {**persons=*2*, recommendation=*unacc***}

We consider all possible splits of this itemset into antecedent and consequent:

- (100.00% ; $\frac{3}{3}$) {**persons**=*2*} ⇒ {**recommendation**=*unacc*}
- (50.00% ; $\frac{3}{6}$) {**recommendation**=*unacc*} ⇒ {**persons**=*2*}

We filter out those rules below our preset confidence requirement (90.00%) and we are left with the following rules that are added to our ruleset:

- (100.00% ; $\frac{3}{3}$) {**persons**=*2*} ⇒ {**recommendation**=*unacc*}

2.2.9 Rules from itemset {**persons=*4*, safety=*med***}

We consider all possible splits of this itemset into antecedent and consequent:

- (50.00% ; $\frac{3}{6}$) {**persons**=*4*} ⇒ {**safety**=*med*}
- (50.00% ; $\frac{3}{6}$) {**safety**=*med*} ⇒ {**persons**=*4*}

None of these have the sufficient confidence (90.00%) so this itemset does not yield any association rules.

2.2.10 Rules from itemset {**safety=*med*, recommendation=*acc***}

We consider all possible splits of this itemset into antecedent and consequent:

- (66.67% ; $\frac{4}{6}$) {**safety**=*med*} ⇒ {**recommendation**=*acc*}
- (100.00% ; $\frac{4}{4}$) {**recommendation**=*acc*} ⇒ {**safety**=*med*}

We filter out those rules below our preset confidence requirement (90.00%) and we are left with the following rules that are added to our ruleset:

- (100.00% ; $\frac{4}{4}$) {**recommendation**=*acc*} ⇒ {**safety**=*med*}

2.2.11 Rules from itemset {**buying-price=*med*, safety=*med*, recommendation=*acc***}

We consider all possible splits of this itemset into antecedent and consequent:

- (100.00% ; $\frac{3}{3}$) {**buying-price**=*med*, **safety**=*med*} ⇒ {**recommendation**=*acc*}
- (100.00% ; $\frac{3}{3}$) {**buying-price**=*med*, **recommendation**=*acc*} ⇒ {**safety**=*med*}
- (75.00% ; $\frac{3}{4}$) {**safety**=*med*, **recommendation**=*acc*} ⇒ {**buying-price**=*med*}
- (75.00% ; $\frac{3}{4}$) {**buying-price**=*med*} ⇒ {**safety**=*med*, **recommendation**=*acc*}
- (50.00% ; $\frac{3}{6}$) {**safety**=*med*} ⇒ {**buying-price**=*med*, **recommendation**=*acc*}
- (75.00% ; $\frac{3}{4}$) {**recommendation**=*acc*} ⇒ {**buying-price**=*med*, **safety**=*med*}

We filter out those rules below our preset confidence requirement (90.00%) and we are left with the following rules that are added to our ruleset:

- (100.00% ; $\frac{3}{3}$) {**buying-price**=*med*, **safety**=*med*} ⇒ {**recommendation**=*acc*}
- (100.00% ; $\frac{3}{3}$) {**buying-price**=*med*, **recommendation**=*acc*} ⇒ {**safety**=*med*}

2.2.12 Rules from itemset {**buying-price=*low*, persons=*2*, recommendation=*unacc***}

We consider all possible splits of this itemset into antecedent and consequent:

- (100.00% ; $\frac{3}{3}$) {**buying-price**=*low*, **persons**=*2*} ⇒ {**recommendation**=*unacc*}
- (100.00% ; $\frac{3}{3}$) {**buying-price**=*low*, **recommendation**=*unacc*} ⇒ {**persons**=*2*}
- (100.00% ; $\frac{3}{3}$) {**persons**=*2*, **recommendation**=*unacc*} ⇒ {**buying-price**=*low*}
- (60.00% ; $\frac{3}{5}$) {**buying-price**=*low*} ⇒ {**persons**=*2*, **recommendation**=*unacc*}
- (100.00% ; $\frac{3}{3}$) {**persons**=*2*} ⇒ {**buying-price**=*low*, **recommendation**=*unacc*}
- (50.00% ; $\frac{3}{6}$) {**recommendation**=*unacc*} ⇒ {**buying-price**=*low*, **persons**=*2*}

We filter out those rules below our preset confidence requirement (90.00%) and we are left with the following rules that are added to our ruleset:

- (100.00% ; $\frac{3}{3}$) {**buying-price**=*low*, **persons**=*2*} ⇒ {**recommendation**=*unacc*}
- (100.00% ; $\frac{3}{3}$) {**buying-price**=*low*, **recommendation**=*unacc*} ⇒ {**persons**=*2*}
- (100.00% ; $\frac{3}{3}$) {**persons**=*2*, **recommendation**=*unacc*} ⇒ {**buying-price**=*low*}
- (100.00% ; $\frac{3}{3}$) {**persons**=*2*} ⇒ {**buying-price**=*low*, **recommendation**=*unacc*}

2.3 Final Association Rule List

We have the collected the following associaction rules (we note the confidence and support percentages over our instances):

- (conf = 100.00% , supp = 25.00%) {**persons**=*2*} ⇒ {**buying-price**=*low*}
- (conf = 100.00% , supp = 25.00%) {**persons**=*2*} ⇒ {**recommendation**=*unacc*}
- (conf = 100.00% , supp = 33.33%) {**recommendation**=*acc*} ⇒ {**safety**=*med*}
- (conf = 100.00% , supp = 25.00%) {**buying-price**=*med*, **safety**=*med*} ⇒ {**recommendation**=*acc*}
- (conf = 100.00% , supp = 25.00%) {**buying-price**=*med*, **recommendation**=*acc*} ⇒ {**safety**=*med*}
- (conf = 100.00% , supp = 25.00%) {**buying-price**=*low*, **persons**=*2*} ⇒ {**recommendation**=*unacc*}
- (conf = 100.00% , supp = 25.00%) {**buying-price**=*low*, **recommendation**=*unacc*} ⇒ {**persons**=*2*}
- (conf = 100.00% , supp = 25.00%) {**persons**=*2*, **recommendation**=*unacc*} ⇒ {**buying-price**=*low*}
- (conf = 100.00% , supp = 25.00%) {**persons**=*2*} ⇒ {**buying-price**=*low*, **recommendation**=*unacc*}