

Association Rule Mining: NLTCB Data

NLTCS (national long-term care survey) consists 35,789 elderly people drawn from 1982 to 2004 (every five years). This paper includes 21574 people that each person's disabilities are recorded in sixteen dimensions (1 if the person is disable; 0 if the person is able in activity). The following analysis presents a methodology known as association rule, which is useful for identifying interesting relations hidden in large database.

➤ Distribution of Numbers of Disabilities

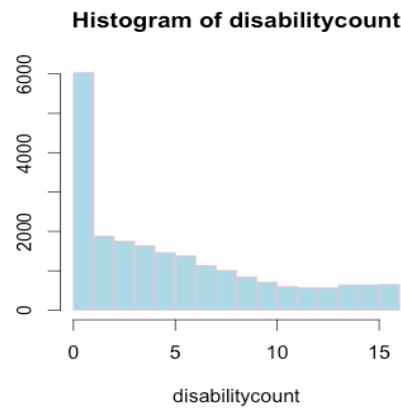


Figure 1. Number of disabilities.

Before commencing analysis, the distribution of the number of disabilities per person shown in Figure 1 provides a check on whether the data have been properly recorded. It can be seen that 3853 (17.85%) people have no disabilities and 660 (0.03%) people are disabled at everything.

➤ Quality Measures

lhs	rhs	support	confidence	lift
33 {Y7,Y11}	=> {Y9}	0.3054139	0.6858541	1.935210
32 {Y7,Y9}	=> {Y11}	0.3054139	0.8833624	1.819000
4 {Y11}	=> {Y9}	0.3114397	0.6413095	1.809523

Figure 2. The top three rules respect to the lift measure.

Mining association rules using the Apriori algorithm that the result is a set of 21215 rules (support > 0.1; confidence > 0.5). Since large sets of rules typically have a large number of different items, we reduce the number of rules by filtering out all rules

with a low support score. The result is a set of 58 rules (support > 0.3; confidence > 0.5). The top three rules are shown in Figure 2. The rule with highest lift is {Y7, Y11} => {Y9}, which indicates that 30.5% elderly people who cannot do heavy housework and shopping, cannot do laundry either. The confidence value demonstrates that amongst all elderly people that have the disability in heavy housework and shopping, 68.5% also have problem in doing laundry. According to lift value, the disability in heavy housework and shopping co-occurs with laundry is 1.93 times more often than they occur independently.

➤ Visualization of Association Rules

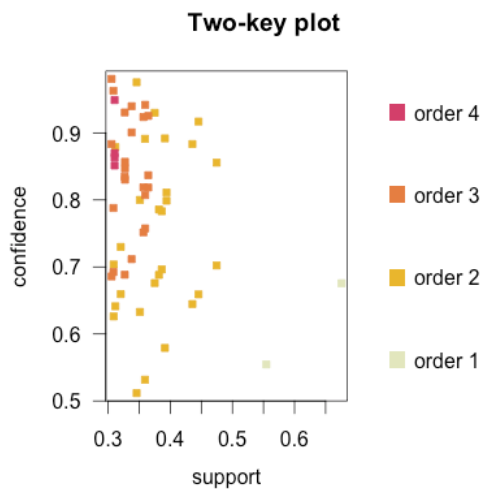


Figure 3. Two-key plot

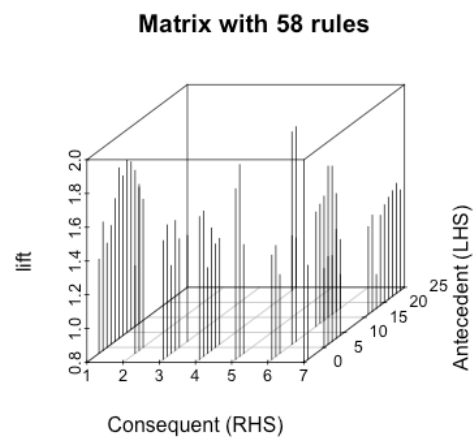


Figure 4. Matrix-based visualization with 3D bars (reordered)

A scatter plot can be used to visualize association rules with two measures on the axes. The results of two-key plot in Figure 3 demonstrate that rules with high lift have a relatively low support, which means order and support have a very strong inverse relationship. The 3D visualization is shown in Figure 4 that organizes the antecedent and consequent items on the x and y-axes respectively. This is useful for exploring different antecedents that have similar impacts on the same consequent in terms of the measure used in the plot.

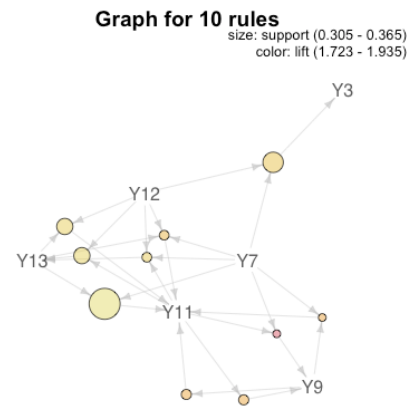
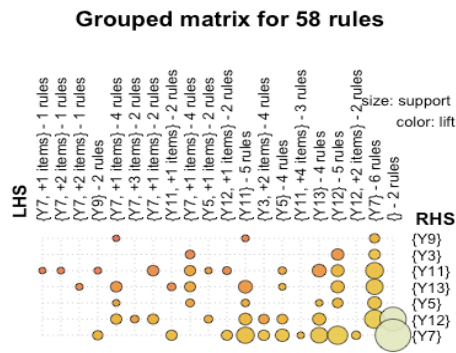


Figure 5. Group matrix-based visualization. Figure 6. Graph-based visualization with items and rules.

The figure 5 provides a balloon plot with antecedent (columns) and consequents (rows). According to lift value, the interesting rules can be found in the middle-west corner. There is 1 rule that contains ‘Y7 (heavy housework)’ and up to 1 other item in the antecedent and the consequent is ‘Y11 (grocery shopping)’. Further, group-based techniques are used in Figure 6, which contains vertices and edges that provide clear representation of items and their relationships.