|  |
| --- |
| > library(arules) > rules.all <- apriori(titanic.raw)  parameter specification: confidence minval smax arem  aval originalSupport support minlen maxlen target   ext         0.8    0.1    1 none FALSE            TRUE     0.1      1     10  rules FALSE  algorithmic control: filter tree heap memopt load sort verbose     0.1 TRUE TRUE  FALSE TRUE    2    TRUE  apriori - find association rules with the apriori algorithm version 4.21 (2004.05.09)        (c) 1996-2004   Christian Borgelt set item appearances ...[0 item(s)] done [0.00s]. set transactions ...[10 item(s), 2201 transaction(s)] done [0.00s]. sorting and recoding items ... [9 item(s)] done [0.00s]. creating transaction tree ... done [0.00s]. checking subsets of size 1 2 3 4 done [0.00s]. writing ... [27 rule(s)] done [0.00s]. creating S4 object  ... done [0.00s]. > rules.all set of 27 rules  > inspect(rules.all)    lhs               rhs             support confidence      lift 1  {}             => {Age=Adult}   0.9504771  0.9504771 1.0000000 2  {Class=2nd}    => {Age=Adult}   0.1185825  0.9157895 0.9635051 3  {Class=1st}    => {Age=Adult}   0.1449341  0.9815385 1.0326798 4  {Sex=Female}   => {Age=Adult}   0.1930940  0.9042553 0.9513700 5  {Class=3rd}    => {Age=Adult}   0.2848705  0.8881020 0.9343750 6  {Survived=Yes} => {Age=Adult}   0.2971377  0.9198312 0.9677574 7  {Class=Crew}   => {Sex=Male}    0.3916402  0.9740113 1.2384742 8  {Class=Crew}   => {Age=Adult}   0.4020900  1.0000000 1.0521033 9  {Survived=No}  => {Sex=Male}    0.6197183  0.9154362 1.1639949 10 {Survived=No}  => {Age=Adult}   0.6533394  0.9651007 1.0153856 11 {Sex=Male}     => {Age=Adult}   0.7573830  0.9630272 1.0132040 12 {Sex=Female,                                                       Survived=Yes} => {Age=Adult}   0.1435711  0.9186047 0.9664669 13 {Class=3rd,                                                        Sex=Male}     => {Survived=No} 0.1917310  0.8274510 1.2222950 14 {Class=3rd,                                                        Survived=No}  => {Age=Adult}   0.2162653  0.9015152 0.9484870 15 {Class=3rd,                                                        Sex=Male}     => {Age=Adult}   0.2099046  0.9058824 0.9530818 16 {Sex=Male,                                                         Survived=Yes} => {Age=Adult}   0.1535666  0.9209809 0.9689670 17 {Class=Crew,                                                       Survived=No}  => {Sex=Male}    0.3044071  0.9955423 1.2658514 18 {Class=Crew,                                                       Survived=No}  => {Age=Adult}   0.3057701  1.0000000 1.0521033 19 {Class=Crew,                                                       Sex=Male}     => {Age=Adult}   0.3916402  1.0000000 1.0521033 20 {Class=Crew,                                                       Age=Adult}    => {Sex=Male}    0.3916402  0.9740113 1.2384742 21 {Sex=Male,                                                         Survived=No}  => {Age=Adult}   0.6038164  0.9743402 1.0251065 22 {Age=Adult,                                                        Survived=No}  => {Sex=Male}    0.6038164  0.9242003 1.1751385 23 {Class=3rd,                                                        Sex=Male,                                                         Survived=No}  => {Age=Adult}   0.1758292  0.9170616 0.9648435 24 {Class=3rd,                                                        Age=Adult,                                                        Survived=No}  => {Sex=Male}    0.1758292  0.8130252 1.0337773 25 {Class=3rd,                                                        Sex=Male,                                                         Age=Adult}    => {Survived=No} 0.1758292  0.8376623 1.2373791 26 {Class=Crew,                                                       Sex=Male,                                                         Survived=No}  => {Age=Adult}   0.3044071  1.0000000 1.0521033 27 {Class=Crew,                                                       Age=Adult,                                                        Survived=No}  => {Sex=Male}    0.3044071  0.9955423 1.2658514 |

通过上面得到的规则有很多是没有意义或者不感兴趣的。假如我们只对右边rhs是survival特征感兴趣，那么通过在appearance中设置rhs=c("Survived=No", "Survived=Yes")即可，另外default="lhs"，其他所有的项将会出现在左边lhs。  
 上面的得到的规则集中第一个规则的lhs是空，这种情况可以通过设置minlen=2来排除。最后通过lift将高可靠度的规则置前。

|  |
| --- |
| > rules <- apriori(titanic.raw, control = list(verbose=F), + parameter = list(minlen=2, supp=0.005, conf=0.8), + appearance = list(rhs=c("Survived=No", "Survived=Yes"), + default="lhs")) > quality(rules) <- round(quality(rules), digits=3) > rules.sorted <- sort(rules, by="lift") > inspect(rules.sorted)    lhs             rhs            support confidence  lift 1  {Class=2nd,                                                 Age=Child}  => {Survived=Yes}   0.011      1.000 3.096 2  {Class=2nd,                                                 Sex=Female,                                                Age=Child}  => {Survived=Yes}   0.006      1.000 3.096 3  {Class=1st,                                                 Sex=Female} => {Survived=Yes}   0.064      0.972 3.010 4  {Class=1st,                                                 Sex=Female,                                                Age=Adult}  => {Survived=Yes}   0.064      0.972 3.010 5  {Class=2nd,                                                 Sex=Female} => {Survived=Yes}   0.042      0.877 2.716 6  {Class=Crew,                                                Sex=Female} => {Survived=Yes}   0.009      0.870 2.692 7  {Class=Crew,                                                Sex=Female,                                                Age=Adult}  => {Survived=Yes}   0.009      0.870 2.692 8  {Class=2nd,                                                 Sex=Female,                                                Age=Adult}  => {Survived=Yes}   0.036      0.860 2.663 9  {Class=2nd,                                                 Sex=Male,                                                  Age=Adult}  => {Survived=No}    0.070      0.917 1.354 10 {Class=2nd,                                                 Sex=Male}   => {Survived=No}    0.070      0.860 1.271 11 {Class=3rd,                                                 Sex=Male,                                                  Age=Adult}  => {Survived=No}    0.176      0.838 1.237 12 {Class=3rd,                                                 Sex=Male}   => {Survived=No}    0.192      0.827 1.222 |

·消除冗余

上面产生的规则中，rule2是rule1的子集，没有提供更多的知识，这种情况下rule2可考虑消除。

|  |
| --- |
| > subset.matrix <- is.subset(rules.sorted, rules.sorted) > subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA > redundant <- colSums(subset.matrix, na.rm=T) >= 1 > which(redundant) [1] 2 4 7 8 > rules.pruned <- rules.sorted[!redundant] > inspect(rules.pruned)   lhs             rhs            support confidence  lift 1 {Class=2nd,                                                Age=Child}  => {Survived=Yes}   0.011      1.000 3.096 2 {Class=1st,                                                Sex=Female} => {Survived=Yes}   0.064      0.972 3.010 3 {Class=2nd,                                                Sex=Female} => {Survived=Yes}   0.042      0.877 2.716 4 {Class=Crew,                                               Sex=Female} => {Survived=Yes}   0.009      0.870 2.692 5 {Class=2nd,                                                Sex=Male,                                                 Age=Adult}  => {Survived=No}    0.070      0.917 1.354 6 {Class=2nd,                                                Sex=Male}   => {Survived=No}    0.070      0.860 1.271 7 {Class=3rd,                                                Sex=Male,                                                 Age=Adult}  => {Survived=No}    0.176      0.838 1.237 8 {Class=3rd,                                                Sex=Male}   => {Survived=No}    0.192      0.827 1.222 |

·解释规则

比如像上面的第一个规则{Class=2nd,  Age=Child}  => {Survived=Yes}，具有高可信度1和可靠度3，而并没有class 1st 和3rd 孩子的幸存记录，这种情况是否说明了class 2nd比其他有更高的存活率呢？不是，因为根本就没有提供相关的数据用于比较。为了这个话题，我们将lhs设置为仅包含"Class=1st", "Class=2nd","Class=3rd", "Age=Child"和"Age=Adult"，rhs是"Survived=Yes"。为了获取更多的记录，我们降低支持度和可信度参数。

|  |
| --- |
| > rules <- apriori(titanic.raw, + parameter = list(minlen=3, supp=0.002, conf=0.2), + appearance = list(rhs=c("Survived=Yes"), + lhs=c("Class=1st", "Class=2nd", "Class=3rd", + "Age=Child", "Age=Adult"), + default="none"), + control = list(verbose=F)) > rules.sorted <- sort(rules, by="confidence") > inspect(rules.sorted)   lhs            rhs                support confidence      lift 1 {Class=2nd,                                                       Age=Child} => {Survived=Yes} 0.010904134  1.0000000 3.0956399 2 {Class=1st,                                                       Age=Child} => {Survived=Yes} 0.002726034  1.0000000 3.0956399 3 {Class=1st,                                                       Age=Adult} => {Survived=Yes} 0.089504771  0.6175549 1.9117275 4 {Class=2nd,                                                       Age=Adult} => {Survived=Yes} 0.042707860  0.3601533 1.1149048 5 {Class=3rd,                                                       Age=Child} => {Survived=Yes} 0.012267151  0.3417722 1.0580035 6 {Class=3rd,                                                       Age=Adult} => {Survived=Yes} 0.068605179  0.2408293 0.7455209 |

可见，头等舱和二等舱的儿童具有相同的幸存率。而三等舱的幸存率明显较低。