Standardising the Lift of an Association Rule

P.D. McNicholas\textsuperscript{a,1,*}, T.B. Murphy\textsuperscript{a,1}, M. O’Regan\textsuperscript{a}

\textit{aDepartment of Statistics, Trinity College Dublin, Ireland.}

Abstract

The lift of an association rule is frequently used, both in itself and as a component in formulae, to gauge the interestingness of a rule. The range of values that lift may take is used to standardise lift so that it is more effective as a measure of interestingness. This standardisation is extended to account for minimum support and confidence thresholds. A method of visualising standardised lift, through the relationship between lift and its upper and lower bounds, is proposed. The application of standardised lift a measure of interestingness is demonstrated on college application data and questionnaire data. In the latter case, negations are introduced into the mining paradigm and an argument for this inclusion is put forward. This argument includes a quantification of the number of extra rules that arise when negations are considered.

\textit{Key words:} Association rules, lift, standardisation, standardised lift, interestingness, college application, Central Applications Office, social life feelings, negative association rules, negations.

1 Association Rules

1.1 Background

Association rules (Agrawal et al. 1993) are used to discover relationships between variables in transaction databases. Analyses based on association rule mining have been conducted on a wide variety of datasets and are particularly useful in the analysis of large datasets.

* Corresponding author.

\textit{Email addresses:} mcnichop@tcd.ie (P.D. McNicholas), murphybt@tcd.ie (T.B. Murphy), moregan@tcd.ie (M. O’Regan).

\textsuperscript{1} This research was funded by an SFI Basic Research Grant (04/BR/M0057).
1.2 Definition of an Association Rule

Given a non-empty set, \( I \), an association rule is a statement of the form \( A \Rightarrow B \), where \( A, B \subset I \) such that \( A \neq \emptyset, B \neq \emptyset \), and \( A \cap B = \emptyset \). The set \( A \) is called the antecedent of the rule, the set \( B \) is called the consequent of the rule, and \( I \) is called the itemset. Association rules are mined over set of transactions, denoted \( \tau = \{ \tau_1, \tau_2, \ldots, \tau_n \} \).

The interestingness of an association rule was originally characterised by functions called ‘support’, ‘confidence’ and ‘lift’. Many other types of interestingness have since been proposed — see Silbershatz & Tuzhilin (1995), Gray & Orlowska (1998) and Dong & Li (1998) for examples.

1.3 Support, Confidence & Lift

The notation \( P(A), P(A, B) \) and \( P(B \mid A) \) is used in the usual way and the afore-mentioned functions are defined in Table 1.

<table>
<thead>
<tr>
<th>Function</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Support</strong></td>
<td>( s(A \Rightarrow B) = P(A, B) ) and ( s(A) = P(A) )</td>
</tr>
<tr>
<td><strong>Confidence</strong></td>
<td>( c(A \Rightarrow B) = P(B \mid A) )</td>
</tr>
<tr>
<td><strong>Expected Confidence</strong></td>
<td>( EC(A \Rightarrow B) = P(B) )</td>
</tr>
<tr>
<td><strong>Lift</strong></td>
<td>( L(A \Rightarrow B) = c(A \Rightarrow B)/P(B) = P(A, B)/(P(A)P(B)) )</td>
</tr>
</tbody>
</table>

Lift represents a measure of the distance between \( P(B \mid A) \) and \( P(B) \), or equivalently, the extent to which \( A \) and \( B \) are not independent. In other words, lift gives a measure of the extent to which the equation

\[ P(A, B) = P(A)P(B) \]

is not true. The expected confidence is the value that confidence would take if \( A \) and \( B \) were in fact independent.
2 Lift

2.1 Range of Values

2.1.1 Range in Terms of $P(A) \& P(B)$

The range of values that the lift of an association rule $A \Rightarrow B$ can take is restricted by the respective values of $P(A)$ and $P(B)$;

$$\frac{\max\{P(A) + P(B) - 1, 1/n\}}{P(A)P(B)} \leq L(A \Rightarrow B) \leq \frac{1}{\max\{P(A), P(B)\}},$$

(1)

where $n$ is the the number of transactions, $\tau_i$. Equation 1 is derived in Appendix A and gives bounds that are almost identical to those derived by Fréchet (1951).

2.1.2 Range in Terms of the Minimum Support Threshold or the Number of Transactions

Equation 2, which is derived in Appendix A, is an expression for the range of values of lift in terms of minimum support threshold, $s$.

$$\frac{4s}{(1 + s)^2} \leq L(A \Rightarrow B) \leq \frac{1}{s}.$$  

(2)

Now, substituting $s = 1/n$ in Equation 2 gives the bounds in terms of $n$,

$$\frac{4n}{(n + 1)^2} \leq L(A \Rightarrow B) \leq n,$$

(3)

which can be written as

$$\frac{4}{n + 1} < L(A \Rightarrow B) \leq n.$$  

(4)

Although straightforward, details of the

also given in Appendix A. These two equations will give realistic bounds for lift in situations where minimum support and confidence thresholds are not used in the mining process.

Standardisation will not be useful in general if achieved through equations 2, 3 or 4 because the limits given in these equations do not vary from rule to rule. In other words, a ranking of rules by lift would be invariant under any standardisation involving these limits alone.
Equations 3 and 4 do, however, give an insight into why it is necessary to standardise lift: the maximum value that lift can take is $n$ and this occurs when a rule is supported by only one transaction. Rules supported by only one transaction will not usually be of interest and may even be pruned in the rule generation stage of the mining process.

2.1.3 Range in Terms of $P(A)$, $P(B)$, Support & Confidence Thresholds

Equation 1 can be expanded to account for both the minimum values of support ($s$) and confidence ($c$) that may be set at the mining stage;

$$\max \left\{ \frac{P(A) + P(B) - 1}{P(A)P(B)}, \frac{4s}{(1 + s)^2}, \frac{1}{P(A)P(B)} \right\} \leq L(A \Rightarrow B) \leq \frac{1}{\max\{P(A), P(B)\}}.$$  \hspace{1cm} (5)

Equation 5 is derived in Appendix A and follows from equations 1 and 2 with little extra work. This equation provides the most general, and accurate, upper and lower bounds for lift given herein.

2.2 Standardisation

Equation 1 can be used to standardise the lift of an association rule according to the formula

$$\mathcal{L}(A \Rightarrow B) = \frac{L(A \Rightarrow B) - \lambda}{v - \lambda},$$  \hspace{1cm} (6)

where

$$\lambda = \frac{\max\{P(A) + P(B) - 1, 1/n\}}{(P(A)P(B))},$$

and

$$v = \frac{1}{\max\{P(A), P(B)\}},$$

are the limits given in Equation 1.

The standardised lift, $\mathcal{L}$, must take a value inside the closed interval $[0, 1]$. A value close to 1 is indicative of an interesting rule. By refining $\lambda$ to account for the lower bound given in Equation 5, $\mathcal{L}$ can be altered to account for minimum support and confidence thresholds. We shall denote this type of standardised lift by $\mathcal{L}^*$, which is given by

$$\mathcal{L}^*(A \Rightarrow B) = \frac{L(A \Rightarrow B) - \lambda^*}{v - \lambda^*},$$  \hspace{1cm} (7)
where
\[
\lambda^* = \max \left\{ \frac{P(A) + P(B) - 1}{P(A)P(B)} \cdot \frac{4s}{(1 + s)^2}, \frac{s}{P(A)P(B)} \cdot \frac{c}{P(B)} \right\}
\] (8)

and \(\nu\) is defined as before.

3 Example I: College Application Data

3.1 Background \& Data

McNicholas (2007) conducted an analysis of central applications office (CAO) data from the year 2000. These data are based on 53,757 applicants who chose up to ten of 533 degree courses. Association rules were initially generated with minimum support set at 0.5\% and a minimum confidence of 80\%. This approach led to the generation of 145 association rules, which were pruned in the following way.

- The constituent items of an association rule were considered.
- If another rule contained these same items, plus at least one other item, then the smaller association rule was deleted.

3.2 Initial Analysis

3.2.1 Rules

Following the application of the pruning procedure mentioned in Section 3.1, 72 of the 145 rules that were initially mined remained. Four of these rules contained only the medicine courses offered across five institutions\(^2\) and they are given, ranked by confidence, in Table 2.

Table 2
The four mined rules that contained solely medicine courses.

<table>
<thead>
<tr>
<th>Id.</th>
<th>Rule</th>
<th>Support</th>
<th>Confidence</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{GY501, CK701, TR051, RC001} (\Rightarrow) {DN002}</td>
<td>0.53%</td>
<td>97.92%</td>
<td>33.76</td>
</tr>
<tr>
<td>2</td>
<td>{CK701, TR051, DN002, RC001} (\Rightarrow) {GY501}</td>
<td>0.53%</td>
<td>92.76%</td>
<td>38.84</td>
</tr>
<tr>
<td>3</td>
<td>{GY501, CK701, DN002, RC001} (\Rightarrow) {TR051}</td>
<td>0.53%</td>
<td>89.52%</td>
<td>41.31</td>
</tr>
<tr>
<td>4</td>
<td>{TR051, GY501, DN002, RC001} (\Rightarrow) {CK701}</td>
<td>0.53%</td>
<td>85.20%</td>
<td>46.54</td>
</tr>
</tbody>
</table>

\(^2\) National University of Ireland Galway (GY), University College Cork (CK), University College Dublin (DN), the Royal College of Surgeons in Ireland (RC) and Trinity College Dublin (TR).
Note that all of the rules in Table 2 have equal support since each of the four rules comprise the same items.

### 3.2.2 Standardised Lift

The limits given in Equation 1 were calculated for each of the rules in Table 2 and are given in Table 3.

**Table 3**
The four rules from Table 2, with their bounds for lift.

<table>
<thead>
<tr>
<th>Id.</th>
<th>Rule</th>
<th>Conf.</th>
<th>Lift</th>
<th>$\lambda$</th>
<th>$\lambda^*$</th>
<th>$\upsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{GY501, CK701, TR051, RC001} ⇒ {DN002}</td>
<td>97.92%</td>
<td>33.76</td>
<td>0.12</td>
<td>31.85</td>
<td>34.48</td>
</tr>
<tr>
<td>2</td>
<td>{CK701, TR051, DN002, RC001} ⇒ {GY501}</td>
<td>92.76%</td>
<td>38.84</td>
<td>0.14</td>
<td>36.64</td>
<td>41.87</td>
</tr>
<tr>
<td>3</td>
<td>{GY501, CK701, DN002, RC001} ⇒ {TR051}</td>
<td>89.52%</td>
<td>41.31</td>
<td>0.14</td>
<td>38.97</td>
<td>46.15</td>
</tr>
<tr>
<td>4</td>
<td>{TR051, GY501, DN002, RC001} ⇒ {CK701}</td>
<td>85.20%</td>
<td>46.54</td>
<td>0.16</td>
<td>43.91</td>
<td>54.62</td>
</tr>
</tbody>
</table>

The lift of each rule in Table 3 and its relationship with the upper and lower limits can be visualised as shown in Figure 1 and Figure 2.

![Figure 1](image1)

Fig. 1. The lift of the rules in Table 3 with upper ($\upsilon$) and lower ($\lambda$) bounds.

Figure 1 and Figure 2 confirm that ordering the rules by confidence gives a ranking that is consistent with the position of the lifts with respect to their respective upper and lower bounds. Figure 2 is, however, much better for this purpose, providing much tighter and more revealing lower bounds. These figures, and Figure 2 in particular, illustrate why rules with greater lift are not necessarily better rules.

The lift of each rule in Table 3 was then standardised according to Equation 6 and Equation 7 and their standardised lifts, $\mathcal{L}$ and $\mathcal{L}^*$ respectively, are given in Table 4.
Fig. 2. The lift of the rules in Table 3 with upper ($\nu$) and lower ($\lambda^*$) bounds.

Table 4
Standardised lifts for the four rules listed in Table 2.

<table>
<thead>
<tr>
<th>Id.</th>
<th>Rule</th>
<th>Supp.</th>
<th>Conf.</th>
<th>Lift</th>
<th>$\mathcal{L}$</th>
<th>$\mathcal{L}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{GY501, CK701, TR051, RC001} $\Rightarrow$ {DN002}</td>
<td>0.53%</td>
<td>97.92%</td>
<td>33.76</td>
<td>0.9791</td>
<td>0.727</td>
</tr>
<tr>
<td>2</td>
<td>{CK701, TR051, DN002, RC001} $\Rightarrow$ {GY501}</td>
<td>0.53%</td>
<td>92.76%</td>
<td>38.84</td>
<td>0.9274</td>
<td>0.420</td>
</tr>
<tr>
<td>3</td>
<td>{GY501, CK701, DN002, RC001} $\Rightarrow$ {TR051}</td>
<td>0.53%</td>
<td>89.52%</td>
<td>41.31</td>
<td>0.8949</td>
<td>0.326</td>
</tr>
<tr>
<td>4</td>
<td>{TR051, GY501, DN002, RC001} $\Rightarrow$ {CK701}</td>
<td>0.53%</td>
<td>85.20%</td>
<td>46.54</td>
<td>0.8516</td>
<td>0.246</td>
</tr>
</tbody>
</table>

The values given for $\mathcal{L}^*$ in Table 4 give a better, more discriminatory, range of values for standardised lift than those given for $\mathcal{L}$. Furthermore, this example illustrates how standardised lift may be used to choose the ‘best’ rule from amongst multiple rules that are comprised of the same items.

Using standardised lift to rank association rules has the effect of ranking a rule depending on the relative position of its lift to the maximum and minimum potential values if its lift. This presents a natural and unambiguous method of ranking association rules.

3.3 Analysis of Rules with Consequent ‘Female’

3.3.1 Rules

McNicholas (2007) also considered rules with a gender variable in the consequent, which was constrained to be singleton. In these cases Gray & Orlowska’s Interestingness (Gray & Orlowska 1998), defined in Equation 9, was used to
rank these rules.

\[
\text{Int}(A \Rightarrow B; K, M) = \left[ P(A, B) \left( \frac{P(A)P(B)}{P(A)P(B)} \right)^K - 1 \right] (P(A), P(B))^M. \tag{9}
\]

McNicholas (2007) set \( K = M = 2 \) in the formula in Equation 9. Table 5 gives the three highest ranked rules with consequent ‘female’. The ranking by \( \text{Int}(A \Rightarrow B; 2, 2) \) is different to that by lift or by confidence.

Table 5
The highest ranked, by \( \text{Int}(A \Rightarrow B; 2, 2) \), rules with consequent ‘female’.

<table>
<thead>
<tr>
<th>Id.</th>
<th>Rule</th>
<th>Support</th>
<th>Confidence</th>
<th>Lift</th>
<th>Int.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{Social Care at DIT} ( \Rightarrow ) {F}</td>
<td>4.99%</td>
<td>94.10%</td>
<td>1.72</td>
<td>0.00165</td>
</tr>
<tr>
<td>2</td>
<td>{Early Child. Care &amp; Ed. at DIT} ( \Rightarrow ) {F}</td>
<td>4.49%</td>
<td>98.13%</td>
<td>1.80</td>
<td>0.00140</td>
</tr>
<tr>
<td>3</td>
<td>{BEd at St. Patrick’s} ( \Rightarrow ) {F}</td>
<td>4.01%</td>
<td>87.23%</td>
<td>1.60</td>
<td>0.00098</td>
</tr>
</tbody>
</table>

3.3.2 Standardised Lift

![Graph showing lift against rule id.](image)

Fig. 3. The lift of the rules in Table 3 with upper (\( \nu \)) and lower (\( \lambda^* \)) bounds.

The upper and lower limits of lift were computed along with the standardised lift, \( L^* \), for each rule in Table 5. These results can be seen in Figure 3 and Table 6.

Table 6
Standardised lift for the three highest ranked rules with consequent ‘female’.

<table>
<thead>
<tr>
<th>Id.</th>
<th>Rule</th>
<th>Support</th>
<th>Confidence</th>
<th>Lift</th>
<th>Int.</th>
<th>( L^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{Social Care at DIT} ( \Rightarrow ) {F}</td>
<td>4.99%</td>
<td>94.10%</td>
<td>1.72</td>
<td>0.00165</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>{Early Child. Care &amp; Ed. at DIT} ( \Rightarrow ) {F}</td>
<td>4.49%</td>
<td>98.13%</td>
<td>1.80</td>
<td>0.00140</td>
<td>0.91</td>
</tr>
<tr>
<td>3</td>
<td>{BEd at St. Patrick’s} ( \Rightarrow ) {F}</td>
<td>4.01%</td>
<td>87.23%</td>
<td>1.60</td>
<td>0.00098</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Note that the ranking of the rules by standardised lift here reflects how close the lift of each rule is to the maximum value that it can take given the support
of the antecedent and consequent. Furthermore, it does not agree with that by Gray and Orlowaska’s interestingness, while it does, in this instance, agree with the ranking by confidence.

3.4 Analysis of Rules with Consequent ‘Male’

3.4.1 Rules

Finally, consider the top three rules, ranked by Gray and Orlowka’s interestingness, with consequent ‘male’. These rules are given in Table 7.

Table 7
The three highest ranked, by $\text{Int}(A \Rightarrow B; 2, 2)$, rules with consequent ‘male’.

<table>
<thead>
<tr>
<th>Id.</th>
<th>Rule</th>
<th>Support</th>
<th>Confidence</th>
<th>Lift</th>
<th>Int.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{Engineering at DIT} ⇒ {M}</td>
<td>2.91%</td>
<td>87.12%</td>
<td>1.92</td>
<td>0.00062</td>
</tr>
<tr>
<td>2</td>
<td>{Engineering at TCD} ⇒ {M}</td>
<td>2.56%</td>
<td>80.30%</td>
<td>1.77</td>
<td>0.00044</td>
</tr>
<tr>
<td>3</td>
<td>{Construction Man. at Waterford IT} ⇒ {M}</td>
<td>2.32%</td>
<td>90.09%</td>
<td>1.98</td>
<td>0.00040</td>
</tr>
</tbody>
</table>

3.4.2 Standardised Lift

Fig. 4. The lift of the rules in Table 7 with upper ($\upsilon$) and lower ($\lambda^*$) bounds.

The lift of these rules, along with upper and lower bounds, is depicted in Figure 4, while the standardised lifts are given in Table 8. Standardised lift presents a different ranking of the rules than Gray and Orlowska’s interestingness. Notably, the rule that was ranked third by Gray and Orlowska’s interestingness is ranked first by standardised lift.
Table 8
Standardised lift ($L^*$) for the three rules with consequent ‘male’.

<table>
<thead>
<tr>
<th>Id.</th>
<th>Rule</th>
<th>Supp.</th>
<th>Conf.</th>
<th>Lift</th>
<th>Int.</th>
<th>$L^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{Engineering at DIT} $\Rightarrow$ {M}</td>
<td>2.91%</td>
<td>87.12%</td>
<td>1.92</td>
<td>0.00062</td>
<td>0.356</td>
</tr>
<tr>
<td>2</td>
<td>{Engineering at TCD} $\Rightarrow$ {M}</td>
<td>2.56%</td>
<td>80.30%</td>
<td>1.77</td>
<td>0.00044</td>
<td>0.015</td>
</tr>
<tr>
<td>3</td>
<td>{Con. Man. at Waterford IT} $\Rightarrow$ {M}</td>
<td>2.32%</td>
<td>90.09%</td>
<td>1.98</td>
<td>0.00040</td>
<td>0.505</td>
</tr>
</tbody>
</table>

3.5 Remark

Despite the results of the analyses in this section, ranking by standardised lift and ranking by confidence is not generally the same. It has been the same in the rules that we have considered with consequent ‘female’ or ‘male’ because the lift was bounded by $[c/P(B), 1/P(B)]$ in these cases and so $L^*$ was given by

$$L - \lambda^* = \frac{P(A, B)}{P(A)P(B)} - \frac{c}{P(B)} = \frac{P(B | A) - c}{1 - c} = \frac{c(A \Rightarrow B) - c}{1 - c},$$

which is a linear function of confidence for fixed $c$. The rules in Section 3.2 were also ranked by confidence, but this was coincidence since in each case the lower bound for lift was given by

$$\lambda^* = \frac{s}{P(A)P(B)}.$$  

4 Example II: German Social Life Feelings

4.1 The Data

The data analysed in this section are taken from a study of ‘social life feelings’ that appeared in Scheussler (1982) and Krebs & Schuessler (1987). This study was also analysed by Bartholomew & Schuessler (1991), Bartholomew (1991), Bartholomew et al. (1997), de Menezes & Bartholomew (1996) and Bartholomew & Knott (1999). The data used herein are the the answers given by a sample of 1,490 Germans to the questions below. These are the data that were analysed by Bartholomew & Knott (1999) and furthermore, they were sourced from the website [3] that is attached to this text.

The following questions were asked in this survey; each was answered ‘yes’ or ‘no’.

(1) Anyone can raise his standard of living if he is willing to work at it.

(2) Our country has too many poor people who can to little to raise their standard of living.
(3) Individuals are poor because of the lack of effort on their part.
(4) Poor people could improve their lot if they tried.
(5) Most people have a good deal of freedom in deciding how to live.

Overall, there were 3,224 ‘yes’ answers and 3,227 ‘no’ answers. The distribution of the number of ‘yes’ answers per question is given in Figure 5, which reveals that only questions 2 and 3 had more than 50% ‘yes’ answers.

![Bar chart showing the number of 'yes' answers per question](image)

Fig. 5. The number of ‘yes’ answers given to each question by the sample of surveyed Germans.

### 4.2 Negations & Coding

These data raise an interesting point; the fact that ‘yes’ is coded ‘1’ and ‘no’ is coded ‘0’ can be viewed as arbitrary. Further, had the questions been worded differently, the ‘1’s and ‘0’ could have been be flipped in some or all of the questions.

For convenience, we introduce the term ‘negation’. Let $x \in I$ denote the presence of an item in a transaction $\tau_i$, then we write $\overline{x}$ to denote the absence or ‘negation’ of $x$ from the transaction $\tau_i$.

In this case, the ‘1’s were coded $y_1$, $y_2$, $y_3$, $y_4$ and $y_5$ respectively while the ‘0’s, or negations, were coded $n_1$, $n_2$, $n_3$, $n_4$ and $n_5$ respectively. Then the apriori algorithm was used to mine the rules.

This straightforward method of mining rules containing negations is facilitated by a relatively small data set and would not be practicable in most cases. Further discussion on negations is given in Section 5.
4.3 Resulting Rules

Association rules were generated using the arules (Hahsler et al. 2005) package in R (R Development Core Team 2005) with minimum support set at 20% and minimum confidence at 80%. This approach led to the generation of 38 association rules, the top ten of which, ranked by $\mathcal{L}^*$, are given in Table 9.

Table 9
Top ten rules from the German ‘social life feeling’ data, ranked by $\mathcal{L}^*$.

<table>
<thead>
<tr>
<th>Id.</th>
<th>Rule</th>
<th>Supp.</th>
<th>Conf.</th>
<th>Lift</th>
<th>$\mathcal{L}^*$</th>
<th>LB</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{n3,n4} $\Rightarrow$ {n1}</td>
<td>0.262</td>
<td>0.951</td>
<td>1.095</td>
<td>0.757</td>
<td>0.920</td>
<td>1.151</td>
</tr>
<tr>
<td>2</td>
<td>{n3,n5} $\Rightarrow$ {n1}</td>
<td>0.255</td>
<td>0.945</td>
<td>1.088</td>
<td>0.726</td>
<td>0.920</td>
<td>1.151</td>
</tr>
<tr>
<td>3</td>
<td>{n4,n5} $\Rightarrow$ {n1}</td>
<td>0.446</td>
<td>0.945</td>
<td>1.087</td>
<td>0.723</td>
<td>0.920</td>
<td>1.151</td>
</tr>
<tr>
<td>4</td>
<td>{n3} $\Rightarrow$ {n1}</td>
<td>0.311</td>
<td>0.935</td>
<td>1.076</td>
<td>0.677</td>
<td>0.920</td>
<td>1.151</td>
</tr>
<tr>
<td>5</td>
<td>{n4} $\Rightarrow$ {n1}</td>
<td>0.548</td>
<td>0.935</td>
<td>1.076</td>
<td>0.674</td>
<td>0.920</td>
<td>1.151</td>
</tr>
<tr>
<td>6</td>
<td>{n2,n4} $\Rightarrow$ {n1}</td>
<td>0.225</td>
<td>0.949</td>
<td>1.092</td>
<td>0.673</td>
<td>0.971</td>
<td>1.151</td>
</tr>
<tr>
<td>7</td>
<td>{n4,n5,y2} $\Rightarrow$ {n1}</td>
<td>0.256</td>
<td>0.934</td>
<td>1.075</td>
<td>0.670</td>
<td>0.920</td>
<td>1.151</td>
</tr>
<tr>
<td>8</td>
<td>{n3,n4,n5} $\Rightarrow$ {n1}</td>
<td>0.221</td>
<td>0.954</td>
<td>1.097</td>
<td>0.667</td>
<td>0.991</td>
<td>1.151</td>
</tr>
<tr>
<td>9</td>
<td>{n4,y2} $\Rightarrow$ {n1}</td>
<td>0.323</td>
<td>0.925</td>
<td>1.064</td>
<td>0.626</td>
<td>0.920</td>
<td>1.151</td>
</tr>
<tr>
<td>10</td>
<td>{n4,n5,y3} $\Rightarrow$ {n1}</td>
<td>0.225</td>
<td>0.936</td>
<td>1.077</td>
<td>0.617</td>
<td>0.958</td>
<td>1.151</td>
</tr>
</tbody>
</table>

All of the rules in Table 9 had the same upper bound for lift because they all had the same consequent, which was bigger in each case, than the antecedent.

The lift of each rule can be seen in context of its upper and lower bounds in Figure 6.

![Lift vs. Rule Id.](image)

Fig. 6. The lift of the ten rules, given in Table 9, along with their upper ($\nu$) lower ($\lambda^*$) bounds.
4.4 Results

Although the rules had not been pruned to any great extent, looking at Table 9, some interesting rules became apparent. The rule with the highest value of standardised lift was \( \{n3, n4\} \Rightarrow \{n1\} \). That is, 95.1% of those who did not agree that people were poor because of lack of effort or that poor people could improve their lot if they tried also did not agree that people could raise their standard of living if they were willing to work at it.

Of course, a more detailed analysis of the results would follow after pruning, perhaps via the pruning method described in Section 3.1. However, these data have been analysed many times, as mentioned in Section 4.1, and are only used here to illustrate standardised lift, \( L^* \).

The ranking of the rules in Table 9 is by \( L^* \) and is different than the ranking by either confidence or lift. This example illustrates \( L^* \) as a new and useful method of ranking association rules. Furthermore, if the pruning described in Section 3.1 was applied to these data, \( L^* \) could be used to effectively break ties.

5 Negations

5.1 Background — Negative Association Rules

Silverstien et al. (1998) introduced the concept of association rules involving negations. Savasere et al. (1998) demonstrated a method of mining rules of the form \( A \Rightarrow \overline{B} \), which they refer to as ‘negative association rules’. Hofmann & Wilhelm (2001) proposed mosaic plots as a means of visualising all combinations of presence or absence of the antecedent and consequent of an association rule; this is a method of deriving rules containing negations after mining, however it is limited in its affectiveness when the consequent is not singleton. Wu et al. (2002, 2004) discussed rules of the form \( A \Rightarrow \overline{B} \), \( \overline{A} \Rightarrow B \) and \( \overline{A} \Rightarrow B \) and provided algorithms to mine such rules.

5.2 Why Include Negations?

Negations were used in Section 4 to facilitate a full analysis of the German social data. Without negations, this analysis would have been prone to error arising out of arbitrary choice of ‘0’ and ‘1’ for each variable.
In general, it will often be the case that the absence of items from the antecedent and the consequent parts of an association rule may of interest. The absence of items from the antecedent part can be related to the presence or absence of items from the consequent part and vice versa. An example of a rule of this form is \{Sunflower Spread\} $\Rightarrow$ \{not Butter\}, such a rule is sometimes referred to as a ‘replacement rule’.

5.3 How Many More Rules

In Appendix B it is shown that the number of rules that can possibly be generated from an itemset consisting of $n$ items and their negations is given by

$$5^n - 2 (3^n) + 1. \quad (12)$$

If negations are not considered, this number is given by

$$\sum_{i=2}^{n} \binom{n}{i} (2^i - 2) = 3^n - 2 (2^n) + 1, \quad (13)$$

where

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$ 

These bounds, and in particular, results that came about in the process of deriving these bounds are similar to the results of Fréchet (1951).

Now, it is trivial to work out the number of ‘extra’ rules that can be mined when negations are included;

$$5^n - 2 (3^n) + 1 - [3^n - 2 (2^n) + 1] = 5^n - 3^{n+1} + 2^{n+1}. \quad (14)$$

When viewed as a proportion of the total number, the amount of rules that contain negations is given by

$$\frac{5^n - 3^{n+1} + 2^{n+1}}{5^n - 2 (3^n) + 1}. \quad (15)$$

Therefore, when an itemset of just 20 items is considered, it follows from Equation 15 that 99.996% of potential association rules involve negations.

Of course, it may be argued that the number of rules given by Equation 12 is unrealistic because a minimum of $2^n$ transactions would be required in order that all potential rules may exist. Furthermore, it is highly unlikely in most practical applications that a transaction involving all, or even most, of the items will occur. For example, in a convenience store application it is highly unlikely that any one customer will purchase more than 0.5% of the items.
However, regardless of these considerations, Equation 15 provides a realistic figure for the proportion of potential rules that will contain negations.

5.4 Mining Association Rules Involving Negations

The huge amount of extra rules that can be mined when negations are included in the mining process presents many computational problems. Furthermore, the method of mining association rules involving negations demonstrated in Section 4 is not viable in general. Thiruvady & Webb (2004) and Gan et al. (2005) presented efficient algorithms to mine rules containing negations.

6 Summary

A new function for ranking association rules, standardised lift, has been introduced. A method of visualising standardised lift or, more precisely, lift relative to its upper and lower bounds, was also introduced. This function, and all theory around it, calls only upon support, confidence and lift — all of which were defined along with the original definition of an association rule (Agrawal et al. 1993) — and, if applicable, minimum support and confidence thresholds.

This new method of ranking is illustrated on two data sets and compared to existing functions of interestingness. In the analysis of second of these datasets, the German social life feeling data, negations were introduced to facilitate the mining process. An argument was given for the inclusion of negations in the association rule mining process, including a quantification of the amount of rules that can be mined when negations are introduced.

References


Dong, G. & Li, J. (1998), Interestingness of discovered association rules in terms of neighbourhood-based unexpectedness, *in ‘Pacific-Asia Conference on Knowledge Discovery and Data Mining’*.


R Development Core Team (2005), R: *A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.


Wu, X., Zhang, C. & Zhang, S. (2002), Mining both positive and negative asso-

A  Lift Theorems

**Theorem A.1** Suppose $I$ is an item set, and $A \Rightarrow B$ is an association rule on a set of transactions $\{\tau_1, \ldots, \tau_n\}$ over $I$. Then

$$\frac{\max\{P(A) + P(B) - 1, 1/n\}}{P(A)P(B)} \leq L(A \Rightarrow B) \leq \frac{1}{\max\{P(A), P(B)\}}.$$

**PROOF.** First consider the bounds of $P(A, B)$. If $P(A) + P(B) < 1$, then the appropriate lower bound is given by

$$P(A, B) \geq \frac{1}{n},$$

otherwise the appropriate lower bound is $P(A, B) \geq P(A) + P(B) - 1$. The upper bound is given by $P(A, B) \leq \min\{P(A), P(B)\}$. Therefore,

$$\max\{P(A) + P(B) - 1, 1/n\} \leq P(A, B) \leq \min\{P(A), P(B)\};$$

and so,

$$\frac{\max\{P(A) + P(B) - 1, 1/n\}}{P(A)P(B)} \leq L(A \Rightarrow B) \leq \frac{1}{\max\{P(A), P(B)\}}.$$

\hfill \Box

**Theorem A.2** Suppose $I$ is an item set, and $A \Rightarrow B$ is an association rule mined over a set of transactions $\{\tau_1, \ldots, \tau_n\}$ over $I$, using support threshold $s$. Then

$$\frac{4s}{(1 + s)^2} \leq L(A \Rightarrow B) \leq \frac{1}{s}.$$

**PROOF.** Let $s = \vartheta/n$. The maximum value of lift will occur when

$$P(A, B) = P(A) = P(B) = \frac{\vartheta}{n}$$

17
and so
\[ L(A \Rightarrow B) \leq \frac{\vartheta/n}{(\vartheta/n)^2} = \frac{n}{\vartheta} = \frac{1}{s}. \]
To find the minimum value of lift, first consider the case where \( n + \vartheta \) is even, then \( L(A \Rightarrow B) \) is minimised when
\[ P(A, B) = \frac{\vartheta}{n} \quad \text{and} \quad P(A) = P(B) = \frac{(n + \vartheta)/2}{n} = \frac{(n + \vartheta)}{2n}, \]
however, if \( n + \vartheta \) is odd then \( L(A \Rightarrow B) \) is minimised when
\[ P(A, B) = \frac{\vartheta}{n}, \quad P(A) = \frac{(n + \vartheta + 1)/2}{n} = \frac{(n + \vartheta + 1)}{2n} \quad \text{and} \quad P(B) = \frac{(n + \vartheta - 1)/2}{n} = \frac{(n + \vartheta - 1)}{2n}. \]
Now,
\[ \frac{(n + \vartheta - 1)}{2n} \cdot \frac{(n + \vartheta + 1)}{2n} = \frac{(n + \vartheta + 1)(n + \vartheta - 1)}{4n^2} < \frac{(n + \vartheta)^2}{4n^2}, \]
therefore,
\[ L(A \Rightarrow B) \geq \frac{\vartheta/n}{(n + \vartheta)^2/4n^2} = \frac{4n\vartheta}{(n + \vartheta)^2} = \frac{4s}{(1 + s)^2}. \]
\[ \square \]

**Corollary A.3** Suppose \( I \) is an item set, and \( A \Rightarrow B \) is an association rule on a set of transactions \( \{\tau_1, \ldots, \tau_n\} \) over \( I \). Then
\[ \frac{4n}{n + 1} < \frac{4n}{(n + 1)^2} \leq L(A \Rightarrow B) \leq n. \]

**PROOF.** Result follows from substituting \( s = 1/n \) into the result of Theorem A.2 and noting that
\[ \frac{4n}{n + 1} < \frac{4n}{(n + 1)^2}. \]
\[ \square \]

**Lemma A.4** Suppose \( I \) is an item set, and \( A \Rightarrow B \) is an association rule on a set of transactions \( \{\tau_1, \ldots, \tau_n\} \) over \( I \). Suppose that minimum thresholds for support and confidence are used in the mining process, denoted \( s \) and \( c \).
respectively. Then
\[
\max \left\{ \frac{P(A) + P(B) - 1}{P(A)P(B)}, \frac{s}{P(A)P(B)}, \frac{c}{P(B)} \right\} \leq L(A \Rightarrow B) \\
\leq \frac{1}{\max\{P(A), P(B)\}}.
\]

**Proof.** Most of the result follows from the results of Theorem A.1 and Theorem A.2. Further, if the argument in Theorem A.2 is repeated with \( s = 1/n \), then it follows that,
\[
\frac{s}{P(A)P(B)} \leq L(A \Rightarrow B).
\]
Then all that remains is to notice that if the minimum confidence threshold is set at \( c \), then
\[
P(B \mid A) \geq c \iff L(A \Rightarrow B) \geq \frac{c}{P(B)}.
\]

\( \square \)

**B Negations Theorem**

**Theorem B.1** The number of rules that can possibly be generated from an itemset consisting of \( n \) items and their negations is given by

\[5^n - 2 \cdot (3^n) + 1.\]

**Proof.** If \( A \) and \( B \) contain a total of \( m \) items then the number of rules involving these \( m \) items and their negations is given by

\[
\left[ \sum_{r=1}^{m-1} mC_r \right] 2^m = \left[ \sum_{r=0}^{m} mC_r - 2 \right] 2^m = (2^m - 2) 2^m = 2^{2m} - 2^{m+1}.
\]

Therefore, from an itemset of size \( n \) there are \( 2^{2n} - 2^{2n+1} \) rules of length \( n \), \( nC_{n-1} \left[ 2^{2(n-1)} - 2^{(n-1)+1} \right] \) rules of length \( n-1 \), \( nC_{n-2} \left[ 2^{2(n-2)} - 2^{(n-2)+1} \right] \) rules of length \( n-2 \) and so on. It follows that the total number of rules that can be generated from these \( n \) items and their negations is given by

\[
\left(2^{2n} - 2^{n+1}\right) + nC_{n-1} \left[ 2^{2(n-1)} - 2^{(n-1)+1} \right] + \ldots + nC_2 \left[ 2^{2(2)} - 2^{2+1} \right].
\]
Which can be expressed as

\[ \sum_{i=2}^{n} nC_i \left( 2^{2i} - 2^{i+1} \right). \]  

(B.1)

Now, the binomial theorem states that

\[(1 + x)^n = \sum_{i=0}^{n} nC_i x^i,\]

and so developing Equation B.1 gives,

\[
\sum_{i=2}^{n} nC_i \left( 2^{2i} - 2^{i+1} \right) = \sum_{i=2}^{n} nC_i 2^{2i} - \sum_{i=2}^{n} nC_i 2^{i+1} = \sum_{i=2}^{n} nC_i 4^i - 2 \sum_{i=2}^{n} nC_i 2^i \\
= \left[ \sum_{i=0}^{n} nC_i 4^i - (nC_0 + nC_1(4)) \right] - 2 \left[ \sum_{i=0}^{n} nC_i 2^i - (nC_0 + nC_1(2)) \right] \\
= \left[ 5^n - (1 + 4n) \right] - 2 \left[ 3^n - (1 + 2n) \right] = 5^n - 2(3^n) + 1.
\]

□