Association rule data mining

Kristian Wichmann

December 31, 2017

1 Transactions and patterns

Let $\mathcal{I}$ be a set of so-called *transactions*. A *pattern* t is a subset of $\mathcal{I}$. So we might consider a dataset $\mathcal{T}$ with n different transactions and m patterns:

$$\mathcal{I} = \{I_1, I_2, \ldots, I_N\}, \qquad \mathcal{T} = \{t_1, t_2, \ldots, t_n\} \tag{1.1}$$

Association rule data mining is an unsupervised learning discipline trying to reveal systematics in the database.

1.1 Support of a pattern

A pattern $X \subseteq \mathcal{I}$ is said to have a *support* equal to the number of elements of $\mathcal{T}$ in which X is a subset. Support may be specified absolutely or relatively:

$$N_X = |X|, \quad \mathrm{supp}X = \frac{X_N}{N} \tag{1.2}$$

2 Association rules

An *association rule* takes the form $X \Rightarrow Y$, where X and Y are disjoint patterns.

For instance, consider:

$$\{\text{beer}\} \Rightarrow \{\text{diapers}\} \tag{2.1}$$

This is an association rule, which says that people who buy beer, typically also purchases diapers. Another (perhaps less surprising) association rule would be:

$$\{\text{cheese, ham, bread}\} \Rightarrow \{\text{butter}\} \tag{2.2}$$

Note the asymmetry between union and intersection here.

2.1 Support of an associaton rule

Like patterns, assocation rules have support associated with them. The support of $X \Rightarrow Y$ is simply equal to the support of $X \cap Y$. Again, this can be specifed absolutely or relatively:

$$N_{X \Rightarrow Y} = N_{X \cup Y} = |X \cap Y|, \quad \mathrm{supp}(X \Rightarrow Y) = \frac{N_{X \cup Y}}{N} \tag{2.3}$$

2.2 Confidence of an associaton rule

The confidence for the association rule $X \Rightarrow Y$ is defined as:

$$\mathrm{conf}(X \Rightarrow Y) = \frac{\mathrm{supp}(X \Rightarrow Y)}{\mathrm{supp}X} = \frac{\mathrm{supp}(X \cup Y)}{\mathrm{supp}X} = \frac{N_{X \Rightarrow Y}}{N_X} = \frac{N_{X \cup Y}}{N_Y} \tag{2.4}$$

So the confidence can be seen as a conditional probability: Given X, what is the chance of Y?

2.3 Lift of an associaton rule

The *lift* for the association rule $X \Rightarrow Y$ is defined as:

$$\mathrm{lift}(X \Rightarrow Y) = \frac{\mathrm{supp}(X \Rightarrow Y)}{\mathrm{supp}X \cdot \mathrm{supp}Y} \tag{2.5}$$

Usng the definition of relative support lift can also be written:

$$\mathrm{lift}(X \Rightarrow Y) = \frac{N_{X \cup Y}/N}{N_X/N \cdot N_Y/N} = \frac{N_{X \cup Y} \cdot N}{N_X \cdot N_Y} \tag{2.6}$$

The interpretation of lift has to do with independence or lack thereof. If lift is equal to 1, this means:

$$1 = \frac{\mathrm{supp}(X \cup Y)}{\mathrm{supp}X \cdot \mathrm{supp}Y} \Leftrightarrow \mathrm{supp}(X \cup Y) = \mathrm{supp}X \cdot \mathrm{supp}Y \tag{2.7}$$

Reinterpreting as probabilities, this means:

$$p(X \cap Y) = p(X) \cdot p(Y) \tag{2.8}$$

In other words, it implies independence between X and Y. Lower values means negative correlation, and higher values positive correlation.

2.4 Conviction of an associaton rule

The *conviction* for the association rule $X \Rightarrow Y$ is defined as:

$$\mathrm{conv}(X \Rightarrow Y) = \frac{1 - \mathrm{supp}Y}{1 - \mathrm{conf}(X \Rightarrow Y)} \tag{2.9}$$