Context: confidence boost implication mining

Task: to identify which closures provide implications having (closure-based) confidence boost above threshold b.

February 19, 2025

Preliminary definitions:

$$mxsup_\tau(X) = \max(\{s(Z) \mid Z = \overline{Z}, Z \supset X, s(Z) \geq \tau\} \cup \{0\})$$

$$mnsub_\tau(X) = \min(\{s(Y) \mid Y = \overline{Y}, Y \subset X, s(Y) \geq \tau\} \cup \{\infty\})$$

In the presence of a support threshold τ, the *support ratio* of an association rule $X \to Y$ is

$$\sigma_\tau(X \to Y) = \frac{s(XY)}{\max\{s(Z) \mid XY \subset Z, s(Z) > \tau\}}$$

We see that this measure does not depend on the antecedent X but just on XY, therefore we may speak of the support ratio of a set Z and understand the support ratio of any of the rules $X \to Y$ with $XY = Z$. We also note that

$$\sigma_\tau(Z) = \frac{s(Z)}{mxsup_\tau(Z)}.$$

1 Confidence Boost

The *confidence boost* of an association rule $X \to Y$ is

$$\beta(X \to Y) = \frac{c(X \to Y)}{\max\{c(X' \to Y') \mid (X \neq X' \vee XY \neq X'Y'), X' \subseteq X, Y \setminus X \subseteq X'Y'\}}.$$

It is clear that for any rule $X \to Y$ with $X \cap Y = Z \neq \emptyset$ we have $s(X \to Y) = s(X \to Y \setminus Z)$, $c(X \to Y) = c(X \to Y \setminus Z)$ and $\beta(X \to Y) = \beta(X \to Y \setminus Z)$, so we do not regard $X \to Y$ and $X \to Y \setminus Z$ as being distinct rules (in particular, if $X \cap Y = \emptyset$, then $X \to XY$ is the same rule as $X \to Y$). In the sequel our rules will always have their antecedents and consequents disjoint. Therefore, condition $Y \setminus X \subseteq X'Y'$ can be written equivalently $Y \subseteq Y'$.

Lemma 1. *If $XY \neq \overline{XY}$ or X is not minimal generator, $\beta(X \to Y) = 1$.*

Proof. Indeed, if XY is not closed, take $X' = X$ and $Y' = \overline{XY} - X \supseteq Y$. It is easy to see that $c(X' \to Y') = c(X \to Y)$.

Moreover, if X is not a minimal generator, take $X' \subset X$ such that $s(X') = s(X)$, and $Y' = XY - X' \supseteq Y$. Clearly, $c(X' \to Y') = c(X \to Y)$. $\square$

Proposition 1. *Consider the rule $X \to Y$ such that $XY = \overline{XY}$, X is a minimal generator for XY and $s(X) = s(XY) > \tau$. Then*

$$\beta(X \to Y) = \begin{cases} \frac{1}{\max\{\frac{mxsup_\tau(\overline{X})}{s(X)}, \max\{\frac{s(X'Y)}{s(X')} \mid X' \subset X\}\}}, & \text{if } \overline{Y} \subset \overline{X} \\ \frac{1}{\max\{\frac{mxsup_\tau(\overline{X})}{s(X)}, \frac{s(X)}{mnsub_\tau(X)}\}}, & \text{if } \overline{Y} = \overline{X} \end{cases}$$

Proof. Since $s(X) = s(XY)$ it is clear that $\overline{X} = \overline{XY}$ (hence, $c(X \to Y) = 1$) and $\overline{Y} \subseteq \overline{X}$. Let us take $X' \to Y'$ to be a distinct rule of maximum confidence that satisfies $X' \subseteq X$ and $Y \subseteq Y'$. Now, if $\overline{Y} = \overline{X}$ then $\overline{X'Y} = \overline{X}$, and one may identify two cases according to whether X' is strictly included in X or not.

- If $X' \subset X$ then $c(X' \to Y') = \frac{s(X'Y')}{s(X')} \leq \frac{s(X'Y)}{s(X')} = \frac{s(X)}{mnsub_\tau(X)}$. Moreover, this upper bound is reached for $Y' = Y$.

- If $X' = X$ then $XY \neq X'Y'$ translates to $XY \subset X'Y'$, and therefore, $c(X' \to Y') = \frac{s(X'Y')}{s(X')} = \frac{mxsup_\tau(XY)}{s(X)} = \frac{mxsup_\tau(\overline{XY})}{s(X)} = \frac{mxsup_\tau(\overline{X})}{s(X)}$.

On the other hand, if $\overline{Y} \subset \overline{X}$ then we can argue as above to conclude that $c(X' \to Y') = \max\{\frac{mxsup_\tau(\overline{X})}{s(X)}, \max\{c(X' \to Y) \mid X' \subset X\}\}$. $\square$

Note that $\beta(X \to \emptyset) = 1$ for all non-empty closed sets X that are their own minimal generators (take $X' \subset X$ and $Y' = \emptyset$).

Next, we show that the set $\mathcal{B}_b = \{X \to Y \mid X \cap Y = \emptyset, XY = \overline{XY}, X \text{ is a minimal generator}, s(X) = s(XY) > \tau, \beta(X \to Y) \geq b\}$ is a GD base for all those implications that pass the support threshold τ and have their confidence boost value greater than or equal to b ($b > 1$).

Indeed, let $X \to Y$ be an arbitrary implication with $s(X \to Y) > \tau$ and $\beta(X \to Y) \geq b$. We can assume without loss of generality that $X \cap Y = \emptyset$. By Lemma 1, $XY = \overline{XY}$ and X is a minimal generator (otherwise, $\beta(X \to Y) = 1 < b$, a contradiction). Hence, $X \to Y$ is in $\mathcal{B}_b$.

On the other hand, none of the rules in $\mathcal{B}_b$ can be obtained from other distinct rule(s) in $\mathcal{B}_b$ by applying one or more times the Armstrong rules:

- augmentation: $\frac{X \to Y}{XX' \to Y}$
  Assume $\mathcal{B}_b$ contains the rules $X \to Y$ and $XX' \to Y$, with $X' \neq \emptyset$. Since $X \to Y$ is a rule (of confidence 1) that participates in computing $\beta(XX' \to Y)$ we get $\beta(XX' \to Y) = 1$, a contradiction.

- transitivity: $\frac{X \to Y, Y \to Z}{X \to Z}$
  Assume $\mathcal{B}_b$ contains the following distinct rules: $X \to Y, Y \to Z$, $X \to Z$. Since XY and XZ are closed sets and $c(X \to Y) = c(X \to Z) = 1$, we

get $XY = XZ$. So, $Y = Z$ (recall that X, Y and Z are pairwise disjoint). It follows $Y = Z = \emptyset$ and therefore the rule $X \to Z$ is the same as $X \to Y$, a contradiction.

- reflexivity: $\frac{\emptyset \to \emptyset}{X \to \emptyset}$
  Assume $\mathcal{B}_b$ contains both $\emptyset \to \emptyset$ and $X \to \emptyset$ for some $X \neq \emptyset$. This means X is a minimal generator for $\overline{X} = X$. Since $X \neq \emptyset$, we can consider an arbitrary $X' \subset X$. The rule $X' \to \emptyset$ is a rule of confidence 1 that participates in computing $\beta(X \to \emptyset)$, so $\beta(X \to \emptyset) = 1$, a contradiction.

The following algorithm outputs the set $\mathcal{B}_b$.

---

set of rules G initially empty
for all Z closed with $s(Z) > \tau$ **do**
  if $\sigma_\tau(Z) \geq b$ **then**
    for X in mingens(Z) **do**
      $Y = Z \backslash X$
      if $\overline{Y} = Z$ and $mnsub_\tau(X)/s(X) \geq b$ **then**
        report $X \to Z$ as output
        add $X \to Y$ to G
      end if
      if $\overline{Y} \subset Z$ and $\min\{s(X')/s(X' \to Y) \mid X' \subset X\} \geq b$ **then**
        report $X \to Z$ as output
        add $X \to Y$ to G
      end if
    end for
  end if
end for

Note that for the case in which $\overline{Y} \subset Z$ we also have

$$\max\{c(X' \to Y) \mid X' \subset X\} \geq \max\{\frac{s(Z)}{s(Z \backslash \overline{Y})}, \frac{s(Y)}{s(\emptyset)}\}$$

(this is easy to see: take $X' = Z \backslash \overline{Y}$ and $X' = \emptyset$, respectively). Unfortunately, it does not help much.

2 Closure Based Confidence Boost

We say that two rules $X_1 \to Y_1$ and $X_2 \to Y_2$ are *equivalent* and we denote it by $(X_1 \to Y_1) \equiv (X_2 \to Y_2)$ if $\overline{X_1} = \overline{X_2}$ and $\overline{X_1 Y_1} = \overline{X_2 Y_2}$.

The *closure-based confidence boost* of an association rule $X \to Y$ is

$$\overline{\beta}(X \to Y) = \frac{c(X \to Y)}{\max\{c(X' \to Y') \mid (\overline{X} \neq \overline{X'} \vee \overline{XY} \neq \overline{X'Y'}), X' \subseteq \overline{X}, Y \backslash X \subseteq \overline{X'Y'}\}}.$$

Clearly, if X and Y are disjoint sets then condition $Y\backslash X \subseteq \overline{X'Y'}$ can be written equivalently $Y \subseteq \overline{X'Y'}$. In the sequel all the rules have the consequent disjoint from the antecedent unless we specify it otherwise.

Proposition 2. *Consider the rule $X \to Y$ with $s(X) = s(XY) > \tau$. Then*

$$\overline{\beta}(X \to Y) = \begin{cases} 1, & \text{if } \overline{Y} \subset \overline{X} \\ \frac{1}{\max\{\frac{mxsup_\tau(\overline{X})}{s(X)}, \frac{s(X)}{mnsub_\tau(\overline{X})}\}}, & \text{if } \overline{Y} = \overline{X} \end{cases}$$

Moreover, $\overline{\beta}(X \to Y) > 1$ if and only if $\overline{Y} = \overline{X}$.

Proof. Since $s(X) = s(XY)$ it is clear that $\overline{X} = \overline{XY}$ (hence, $c(X \to Y) = 1$) and $\overline{Y} \subseteq \overline{X}$. Now, if $\overline{Y} \subset \overline{X}$ then one can take $X' = Y$ and $Y' = \emptyset$. It is easy to check that the rule $X' \to Y'$ is not equivalent with $X \to Y$ and it satisfies $X' \subseteq \overline{X}$ and $Y \subseteq \overline{X'Y'}$. Since $c(X' \to Y') = 1$ we get $\overline{\beta}(X \to Y) = 1$.

On the other hand, if $\overline{Y} = \overline{X}$, let us take $X' \to Y'$ to be a non-equivalent rule of maximum confidence that satisfies $X' \subseteq \overline{X}$ and $Y \subseteq \overline{X'Y'}$ (or, equivalently, that $\overline{X'} \subseteq \overline{X}$ and $\overline{Y} \subseteq \overline{X'Y'}$). We distinguish two cases according to whether $\overline{X'}$ is strictly included in $\overline{X}$ or not.

- If $\overline{X'} \subset \overline{X}$ then $c(X' \to Y') = \frac{s(\overline{X'Y'})}{s(\overline{X'})} \leq \frac{s(Y)}{mnsub_\tau(\overline{X})} = \frac{s(X)}{mnsub_\tau(\overline{X})}$. This maximum value is reached for $Y' = \overline{Y} - X'$, so $c(X' \to Y') = \frac{s(X)}{mnsub_\tau(\overline{X})}$.
- If $\overline{X'} = \overline{X}$ then $\overline{XY} \neq \overline{X'Y'}$ translates to $\overline{XY} \subset \overline{X'Y'}$, and therefore, $c(X' \to Y') = \frac{s(\overline{X'Y'})}{s(\overline{X'})} = \frac{mxsup_\tau(\overline{XY})}{s(\overline{X'})} = \frac{mxsup_\tau(\overline{X})}{s(X)}$.

□

Corollary 1. *In the conditions of Proposition 1, $\beta(X \to Y) \geq \overline{\beta}(X \to Y)$.*

Proof. If $\overline{Y} \subset \overline{X}$, $\overline{\beta}(X \to Y) = 1$ and the inequality trivially holds. Otherwise, $\overline{\beta}(X \to Y) = \frac{1}{\max\{\frac{mxsup_\tau(\overline{X})}{s(X)}, \frac{s(X)}{mnsub_\tau(\overline{X})}\}} \leq \frac{1}{\max\{\frac{mxsup_\tau(\overline{X})}{s(X)}, \frac{s(X)}{mnsub_\tau(X)}\}} = \beta(X \to Y)$ because $mnsub_\tau(\overline{X}) \leq mnsub_\tau(X)$. □

Next, we show that the set $\overline{\mathcal{B}}_b = \{X \to Y \mid X \cap Y = \emptyset, XY = \overline{XY}, X \text{ is a minimal generator}, s(X) = s(XY) > \tau, \overline{\beta}(X \to Y) \geq b\}$ is a GD base for all implications of closure based confidence boost greater than or equal to b ($b > 1$) that pass the support threshold τ.

Lemma 2. *For all implications $X \to Y$ with $\overline{\beta}(X \to Y) \geq b$ and $s(X) > \tau$ it holds that either the rule itself or an equivalent one is in $\overline{\mathcal{B}}_b$.*

Proof. Let $X \to Y$ be a rule with $s(X) = s(XY) > \tau$ and $\overline{\beta}(X \to Y) \geq b$. We show that there exists an equivalent rule $X' \to Y'$ in $\overline{\mathcal{B}}_b$. Indeed, let $X' \subseteq X$ be a minimal generator for $\overline{XY}$ and $Y' = \overline{XY}\backslash X'$.

- since $\overline{X'} \subseteq \overline{X} \subseteq \overline{XY}$ and $s(X') = s(XY)$ we get $\overline{X'} = \overline{X}$.

- $X'Y' = \overline{XY}$ implies $\overline{X'Y'} = \overline{XY} = X'Y'$
- $X' \cap Y' = \emptyset$

So, $X \to Y$ and $X' \to Y'$ are two equivalent rules and $X' \to Y'$ is in $\overline{\mathcal{B}}_b$. $\square$

Lemma 3. *Let $X \to Y$ be a rule with $X \cap Y = \emptyset$ and $Z = XY$. Then $X \to Y$ is in $\overline{\mathcal{B}}_b$ if and only if $Z = \overline{Z}$, X is a minimal generator, $s(Y) = s(X) = s(Z) > \tau$, $\sigma_\tau(Z) \geq b$ and $mnsub_\tau(Z)/s(Z) \geq b$.*

Proof. The "only if" direction is immediate: $s(X) = s(XY)$ implies $\overline{Y} \subseteq \overline{X}$, and since $\overline{\beta}(X \to Y) \geq b > 1$ we get $\overline{Y} = \overline{X}$ (see Proposition 2). Moreover, $\max\{\frac{mxsup_\tau(\overline{X})}{s(X)}, \frac{s(X)}{mnsub_\tau(\overline{X})}\} \leq \frac{1}{b}$, so $\sigma_\tau(Z) \geq b$ and $mnsub_\tau(Z)/s(X) \geq b$ (recall that $Z = XY = \overline{XY} = \overline{X}$).

For the "if" direction, we only have to argue that $\overline{\beta}(X \to Y)$ is indeed not smaller than b. But since $s(Y) = s(X) = s(XY)$ we get $\overline{Y} = \overline{X} = \overline{XY} = Z$, so $\overline{\beta}(X \to Y) = \frac{1}{\max\{\frac{mxsup_\tau(\overline{X})}{s(X)}, \frac{s(X)}{mnsub_\tau(\overline{X})}\}}$ (see Proposition 2). We can therefore conclude that $\overline{\beta}(X \to Y) \geq b$.

$\square$

Lemma 4. *None of the rules in $\overline{\mathcal{B}}_b$ can be obtained from other distinct rule(s) in $\overline{\mathcal{B}}_b$ by applying one or more times the Armstrong rules:*

Proof.

- augmentation: $\frac{X \to Y}{XX' \to Y}$
  If $\overline{X} = \overline{XX'}$, the case is solved ($X$ is minimal generator).
  If $\overline{X} \subset \overline{XX'}$, the rule $X \to Y$ is a rule of confidence 1 that participates in the computation of $\overline{\beta}(XX' \to Y)$, so $\overline{\beta}(XX' \to Y) = 1$.

- transitivity: $\frac{X \to Y, Y \to Z}{X \to Z}$
  The same as in the case of confidence boost:
  Since X is minimal generator for both $\overline{XY}$ and $\overline{XZ}$ (XY and XZ are closed sets), it follows $XY = XZ$ and hence $Y = Z$ (recall that X, Y and Z are pairwise disjoint). We get $Y = Z = \emptyset$ and therefore the rule $X \to Z$ is the same as $X \to Y$.

- reflexivity: $\frac{\emptyset \to \emptyset}{X \to \emptyset}$
  Assume $X \to \emptyset$ is in $\overline{\mathcal{B}}_b$ for some $X \neq \emptyset$. Then X is minimal generator for $\overline{X}$ and $X = \overline{X}$. Since $\emptyset \to \emptyset$ is a rule of confidence 1 that participates in the computation of $\overline{\beta}(X \to \emptyset)$ ($\overline{X} \neq \overline{\emptyset}$ because the minimal generator for $\overline{\emptyset}$ is the empty set itself), we can conclude that $\overline{\beta}(X \to \emptyset) = 1$.

$\square$

The following algorithm outputs the set $\overline{\mathcal{B}}_b$.

set of rules G initially empty
for all Z closed with $s(Z) > \tau$ **do**
  if $\sigma_\tau(Z) \geq b$ and $mnsub_\tau(Z)/s(Z) \geq b$ **then**
    for X in mingens(Z) **do**
      if $\overline{Z \backslash X} = Z$ **then**
        report $X \rightarrow Z$ as output
        add $X \rightarrow Z \backslash X$ to G
      end if
    end for
  end if
end for