

Explanations for Itemset Mining by Constraint Programming: A Case Study using ChEMBL data (Supplementary Material)

Maksim Koptelov¹, Albrecht Zimmermann¹, Patrice Boizumault¹, Ronan Bureau², and Jean-Luc Lamotte¹

¹ UNICAEN, ENSICAEN, CNRS – UMR GREYC, 14000, Caen, France

² CERMN/UNICAEN, 14200, Caen, France

name.lastname@unicaen.fr

Table 1. Refined constraints for itemset mining and their adaptation for a CSP solver

Name	Constraint	Adaptation for a solver
Coverage	$T_i = 1 \leftrightarrow \sum_i (1 - DB_{ti}) \cdot I_i = 0$	$c(t) = \sum_i (1 - DB_{ti}) \cdot I_i^{min} > 0$ \rightarrow remove 1 from T_t $c(t) = \sum_i (1 - DB_{ti}) \cdot I_i^{max} = 0$ \rightarrow remove 0 from T_t
Emergence	$I_i = 1 \rightarrow f\left(\sum_{t+} DB_{ti} \cdot T_t, \sum_{t-} DB_{ti} \cdot T_t\right) \geq \theta_1$	$c(i) = \max\left(f\left(\sum_{t+} DB_{ti} \cdot T_t^{max}, \sum_{t-} DB_{ti} \cdot T_t^{min}\right), f\left(\sum_{t+} DB_{ti} \cdot T_t^{min}, \sum_{t-} DB_{ti} \cdot T_t^{max}\right)\right) \leq \theta_1$ \rightarrow remove 1 from I_i
Closure	$I_i = 1 \leftrightarrow \sum_t (1 - DB_{ti}) \cdot T_t = 0$	$c(i) = \sum_t (1 - DB_{ti}) \cdot T_t^{min} > 0$ \rightarrow remove 1 from I_i $c(i) = \sum_t (1 - DB_{ti}) \cdot T_t^{max} = 0$ \rightarrow remove 0 from I_i
Frequency	$I_i = 1 \rightarrow \sum_t DB_{ti} \cdot T_t \geq \theta_2$	$c(i) = \sum_t DB_{ti} \cdot T_t^{max} < \theta_2$ \rightarrow remove 1 from I_i
Size	$T_i = 1 \rightarrow \sum_i DB_{ti} \cdot I_i \leq \theta_3$	$c(t) = \sum_i DB_{ti} \cdot I_i^{min} > \theta_3$ \rightarrow remove 1 from T_t
Purity	$I_i = 1 \rightarrow \min\left(\sum_{t+} DB_{ti} \cdot T_t, \sum_{t-} DB_{ti} \cdot T_t\right) = 0$	$c(i) = \min\left(\sum_{t+} DB_{ti} \cdot T_t^{min}, \sum_{t-} DB_{ti} \cdot T_t^{min}\right) \neq 0$ \rightarrow remove 1 from I_i

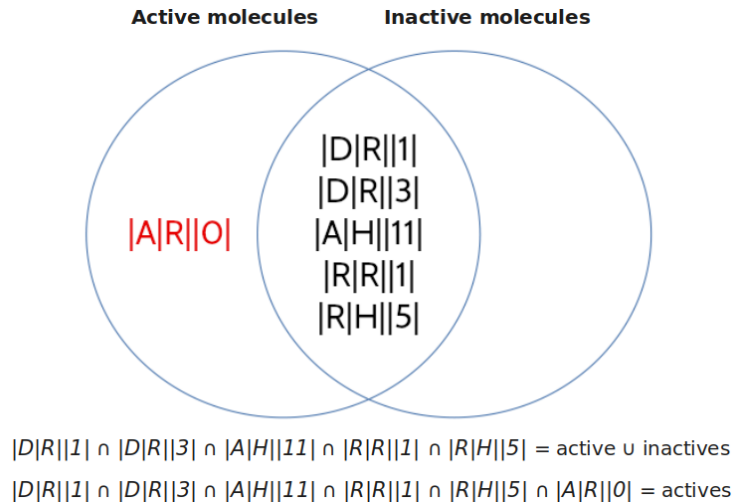


Fig. 1. Active, inactive molecules and their intersection

A APPENDIX

c_{2361} , c_{2372} , c_{2440} , c_{2461} , c_{2488} – purity constraints after filtering the CSP:

$$\begin{aligned}
 c_{2361}: & |D|R||1| = 1 \rightarrow \min(T_1 + T_2 + \dots + T_{356} + T_{379}, T_{429}) = 0 \text{ *False*} \\
 c_{2372}: & |D|R||3| = 1 \rightarrow \min(T_1 + T_2 + \dots + T_{356} + T_{379}, T_{429}) = 0 \text{ *False*} \\
 c_{2440}: & |A|H||11| = 1 \rightarrow \min(T_1 + T_2 + \dots + T_{356} + T_{379}, T_{429}) = 0 \text{ *False*} \\
 c_{2461}: & |R|R||1| = 1 \rightarrow \min(T_1 + T_2 + \dots + T_{356} + T_{379}, T_{429}) = 0 \text{ *False*} \\
 c_{2488}: & |R|H||5| = 1 \rightarrow \min(T_1 + T_2 + \dots + T_{356} + T_{379}, T_{429}) = 0 \text{ *False*}
 \end{aligned}$$

B APPENDIX

if there is

$|P|P||3|$, $|P|D||10|$, $|P|D||11|$, $|P|D||12|$, $|P|D||14|$, $|P|D||17|$, $|P|D||19|$, $|P|D||20|$, $|P|D||3|$, $|P|D||4|$,
 $|P|D||5|$, $|P|D||6|$, $|P|D||7|$, $|P|D||8|$, $|P|D||9|$, $|P|A||10|$, $|P|A||11|$, $|P|A||12|$, $|P|A||13|$, $|P|A||14|$,
 $|P|A||15|$, $|P|A||16|$, $|P|A||17|$, $|P|A||18|$, $|P|A||19|$, $|P|A||20|$, $|P|A||22|$, $|P|A||3|$, $|P|A||4|$, $|P|A||5|$,
 $|P|A||6|$, $|P|A||7|$, $|P|A||8|$, $|P|A||9|$, $|P|R||10|$, $|P|R||11|$, $|P|R||12|$, $|P|R||13|$, $|P|R||14|$, $|P|R||15|$,
 $|P|R||16|$, $|P|R||17|$, $|P|R||18|$, $|P|R||2|$, $|P|R||4|$, $|P|R||6|$, $|P|R||7|$, $|P|R||8|$, $|P|R||9|$, $|P|H||0|$,
 $|P|H||10|$, $|P|H||11|$, $|P|H||12|$, $|P|H||13|$, $|P|H||14|$, $|P|H||15|$, $|P|H||16|$, $|P|H||18|$, $|P|H||19|$, $|P|H||4|$,
 $|P|H||7|$, $|P|H||8|$, $|N|D||0|$, $|N|D||8|$, $|N|D||9|$, $|N|A||0|$, $|N|A||10|$, $|N|A||12|$, $|N|A||8|$, $|N|A||9|$,
 $|N|R||1|$, $|N|R||2|$, $|N|H||9|$, $|D|D||11|$, $|D|D||12|$, $|D|D||13|$, $|D|D||14|$, $|D|D||16|$, $|D|D||2|$, $|D|D||3|$,
 $|D|D||4|$, $|D|D||5|$, $|D|D||6|$, $|D|D||7|$, $|D|D||8|$, $|D|D||9|$, $|D|A||1|$, $|D|A||13|$, $|D|A||14|$, $|D|A||15|$,
 $|D|A||16|$, $|D|A||17|$, $|D|R||10|$, $|D|R||11|$, $|D|R||12|$, $|D|R||13|$, $|D|R||14|$, $|D|R||15|$, $|D|R||16|$,
 $|D|R||18|$, $|D|R||19|$, $|D|H||13|$, $|D|H||14|$, $|D|H||15|$, $|D|H||16|$, $|D|H||17|$, $|D|H||18|$, $|D|H||19|$,
 $|D|H||21|$, $|D|H||3|$, $|A|A||1|$, $|A|A||13|$, $|A|A||14|$, $|A|A||15|$, $|A|A||16|$, $|A|A||17|$, $|A|A||18|$, $|A|R||13|$,
 $|A|R||14|$, $|A|R||15|$, $|A|R||16|$, $|A|R||17|$, $|A|R||18|$, $|A|R||19|$, $|A|H||0|$, $|A|H||16|$, $|A|H||17|$,
 $|A|H||18|$, $|A|H||19|$, $|A|H||20|$, $|A|H||21|$, $|A|H||23|$, $|R|R||10|$, $|R|R||11|$, $|R|R||12|$, $|R|R||13|$,
 $|R|R||6|$, $|R|R||7|$, $|R|H||14|$, $|R|H||15|$, $|R|H||16|$, $|R|H||17|$, $|R|H||18|$, $|R|H||19|$, $|H|H||12|$, $|H|H||13|$,
 $|H|H||14|$, $|H|H||15|$, $|H|H||16|$, $|H|H||17|$, $|H|H||18|$, $|H|H||19|$, $|H|H||2|$, $|H|H||3|$, $|H|H||4|$, $|H|H||6|$,
 $|H|H||7|$, $|H|H||8|$, $|H|H||9|$

in the pattern, it is always not frequent

C APPENDIX

if there is

$|P|P||3|$, $|P|D||10|$, $|P|D||11|$, $|P|D||12|$, $|P|D||14|$, $|P|D||17|$, $|P|D||19|$, $|P|D||20|$, $|P|D||3|$, $|P|D||4|$,
 $|P|D||5|$, $|P|D||6|$, $|P|D||7|$, $|P|D||8|$, $|P|D||9|$, $|P|A||10|$, $|P|A||11|$, $|P|A||12|$, $|P|A||13|$, $|P|A||14|$,
 $|P|A||15|$, $|P|A||16|$, $|P|A||17|$, $|P|A||18|$, $|P|A||19|$, $|P|A||20|$, $|P|A||22|$, $|P|A||3|$, $|P|A||4|$, $|P|A||5|$,
 $|P|A||6|$, $|P|A||7|$, $|P|A||8|$, $|P|A||9|$, $|P|R||10|$, $|P|R||11|$, $|P|R||12|$, $|P|R||13|$, $|P|R||14|$, $|P|R||15|$,
 $|P|R||16|$, $|P|R||17|$, $|P|R||18|$, $|P|R||2|$, $|P|R||4|$, $|P|R||5|$, $|P|R||6|$, $|P|R||7|$, $|P|R||8|$, $|P|R||9|$,
 $|P|H||0|$, $|P|H||1|$, $|P|H||10|$, $|P|H||11|$, $|P|H||12|$, $|P|H||13|$, $|P|H||14|$, $|P|H||15|$, $|P|H||16|$, $|P|H||18|$,
 $|P|H||19|$, $|P|H||4|$, $|P|H||7|$, $|P|H||8|$, $|N|D||0|$, $|N|D||8|$, $|N|D||9|$, $|N|A||0|$, $|N|A||10|$, $|N|A||12|$,
 $|N|A||8|$, $|N|A||9|$, $|N|R||1|$, $|N|R||2|$, $|N|H||9|$, $|D|D||10|$, $|D|D||11|$, $|D|D||12|$, $|D|D||13|$, $|D|D||14|$,
 $|D|D||16|$, $|D|D||2|$, $|D|D||3|$, $|D|D||4|$, $|D|D||5|$, $|D|D||6|$, $|D|D||7|$, $|D|D||8|$, $|D|D||9|$, $|D|A||0|$,
 $|D|A||1|$, $|D|A||10|$, $|D|A||11|$, $|D|A||12|$, $|D|A||13|$, $|D|A||14|$, $|D|A||15|$, $|D|A||16|$, $|D|A||17|$,
 $|D|R||0|$, $|D|R||10|$, $|D|R||11|$, $|D|R||12|$, $|D|R||13|$, $|D|R||14|$, $|D|R||15|$, $|D|R||16|$, $|D|R||18|$,
 $|D|R||19|$, $|D|R||8|$, $|D|R||9|$, $|D|H||1|$, $|D|H||11|$, $|D|H||12|$, $|D|H||13|$, $|D|H||14|$, $|D|H||15|$, $|D|H||16|$,
 $|D|H||17|$, $|D|H||18|$, $|D|H||19|$, $|D|H||21|$, $|D|H||3|$, $|D|H||6|$, $|D|H||7|$, $|D|H||8|$, $|A|A||1|$, $|A|A||10|$,
 $|A|A||11|$, $|A|A||12|$, $|A|A||13|$, $|A|A||14|$, $|A|A||15|$, $|A|A||16|$, $|A|A||17|$, $|A|A||18|$, $|A|R||11|$,
 $|A|R||12|$, $|A|R||13|$, $|A|R||14|$, $|A|R||15|$, $|A|R||16|$, $|A|R||17|$, $|A|R||18|$, $|A|R||19|$, $|A|H||0|$, $|A|H||1|$,
 $|A|H||10|$, $|A|H||13|$, $|A|H||14|$, $|A|H||15|$, $|A|H||16|$, $|A|H||17|$, $|A|H||18|$, $|A|H||19|$, $|A|H||20|$,
 $|A|H||21|$, $|A|H||23|$, $|R|R||10|$, $|R|R||11|$, $|R|R||12|$, $|R|R||13|$, $|R|R||5|$, $|R|R||6|$, $|R|R||7|$, $|R|R||9|$,
 $|R|H||10|$, $|R|H||11|$, $|R|H||12|$, $|R|H||13|$, $|R|H||14|$, $|R|H||15|$, $|R|H||16|$, $|R|H||17|$, $|R|H||18|$,
 $|R|H||19|$, $|H|H||10|$, $|H|H||11|$, $|H|H||12|$, $|H|H||13|$, $|H|H||14|$, $|H|H||15|$, $|H|H||16|$, $|H|H||17|$,
 $|H|H||18|$, $|H|H||19|$, $|H|H||2|$, $|H|H||3|$, $|H|H||4|$, $|H|H||5|$, $|H|H||6|$, $|H|H||7|$, $|H|H||8|$, $|H|H||9|$

in the pattern, it is always not emerged