Appendix

Concise Representations of Frequent Patterns

In the literature, besides the generator representation, the commonly used concise representations of frequent patterns includes:

The Maximal pattern representation. Maximal patterns are first introduced in [3]. Frequent maximal patterns refer to the longest patterns that are frequent.

Definition 4 (Maximal Pattern). *Given a dataset $\mathcal{D}$ and a support threshold ms, a pattern P is a frequent "maximal pattern" iff $sup(P, \mathcal{D}) \geq ms$ and, for every $Q \supset P$, it is the case that $sup(Q, \mathcal{D}) < ms$.*

The maximal pattern representation is composed with the set of frequent maximal patterns enriched with their support values. Based on the *a priori* property of frequent patterns, the maximal pattern representation is able to enumerate all frequent patterns. However, the representation lacks the information to derive the exact support of frequent patterns. The maximal pattern representation is a lossy representation.

The Closed pattern representation. Unlike the maximal pattern representation, the closed pattern representation is a lossless concise representation of frequent patterns.

Definition 5 (Closed Pattern). *Given a dataset $\mathcal{D}$, a pattern P is a "closed pattern" iff for every $P' \supset P$, it is the case that $sup(P', \mathcal{D}) < sup(P, \mathcal{D})$.*

For a dataset $\mathcal{D}$ and support threshold ms, the closed pattern representation is constructed with the set of frequent closed patterns, denoted as $\mathcal{FC}(\mathcal{D}, ms)$, and their corresponding support information. Not only can enumerate all frequent patterns, the closed pattern representation is also able to derive the support of all frequent patterns. For any frequent pattern P in dataset $\mathcal{D}$, its support can be calculated as:

$$sup(P, \mathcal{D}) = max\{sup(C, \mathcal{D}) | C \supseteq P, C \in \mathcal{FC}(\mathcal{D}, ms)\} \quad (2)$$

The Free-set representation. The concept of free-sets is proposed by Boulicaut et. al. in [4] as a condensed representation of patterns.

Definition 6 (δ-free-set). *Given a dataset $\mathcal{D}$, a pattern P is a "δ-free-set" iff, for every pattern $Q \subset P$, it is the case that $sup(Q, \mathcal{D}) - sup(P, \mathcal{D}) > \delta$. The set of all δ-free-sets w.r.t. $\mathcal{D}$ is noted $Free(\mathcal{D}, \delta)$. Given a support threshold ms, a δ-free-set P is a "frequent δ-free-set" iff $sup(P, \mathcal{D}) \geq ms$. The set of all frequent δ-free-sets w.r.t. $\mathcal{D}$ and ms is noted $FreqFree(\mathcal{D}, \delta, ms)$.*

The free-set representation of frequent patterns is composed with the set of frequent δ-free-sets, $FreqFree(\mathcal{D}, \delta, ms)$, the negative border of δ-free-sets and their corresponding support values. The negative border of the δ-free-sets is denoted by $NBd(FreqFree)$ and defined as: $NBd(FreqFree) = \{P|P \notin FreqFree(\mathcal{D}, \delta, ms) \wedge (\forall Q \subset P, Q \in FreqFree(\mathcal{D}, \delta, ms))\}$.

The Disjunctive-free set representation. Inspired by the idea of free-sets, disjunctive-free sets are formally defined as follows.

Definition 7 (k-disjunctive-free set). *Given a dataset $\mathcal{D}$, a pattern P is a "K-disjunctive-free set"" iff there does not exist pattern $Q \subset P$ such that $|P| - |Q| \leq k$ and $sup(P, \mathcal{D}) = \sum_{Q \subseteq Q' \subset P} (-1)^{|P|-|Q'|-1} \cdot sup(Q', \mathcal{D})$. Given a support threshold ms, a k-disjunctive-free set P is a "frequent k-disjunctive-free set" iff $sup(P, \mathcal{D}) \geq ms$. The set of all frequent k-disjunctive-free sets w.r.t. $\mathcal{D}$ and ms is noted $FreqDFree(\mathcal{D}, k, ms)$.*

The disjunctive-free set representation of frequent patterns is composed with the set of frequent k-disjunctive-free sets, $FreqDFree(\mathcal{D}, k, ms)$, the negative border of k-disjunctive-free sets and their support values. The negative border of the k-disjunctive-free sets is denoted by $NBd(FreqDFree)$ and defined as: $NBd(FreqDFree) = \{P|P \notin FreqDFree(\mathcal{D}, k, ms) \wedge (\forall Q \subset P, Q \in FreqDFree(\mathcal{D}, k, ms))\}$.