

Listing closed sets of strongly accessible set systems with applications to data mining*

Mario Boley^{1,2} Tamás Horváth^{1,2} Axel Poigné² Stefan Wrobel^{1,2}

¹Dept. of Computer Science III, University of Bonn, Germany

²Fraunhofer IAIS, D-53754 Sankt Augustin, Germany

{mario.boleym,tamas.horvath,axel.poigne,stefan.wrobel}@iais.fraunhofer.de

Abstract

We study the problem of listing all closed sets of a closure operator σ that is a partial function on the power set of some finite ground set E , i.e., $\sigma : \mathcal{F} \rightarrow \mathcal{F}$ with $\mathcal{F} \subseteq \mathcal{P}(E)$. A very simple divide-and-conquer algorithm is analyzed that correctly solves this problem if and only if the domain of the closure operator is a strongly accessible set system. Strong accessibility is a strict relaxation of greedoids as well as of independence systems. This algorithm turns out to have delay $O(|E|(T_{\mathcal{F}} + T_{\sigma} + |E|))$ and space $O(|E| + S_{\mathcal{F}}S_{\sigma})$, where $T_{\mathcal{F}}$, $S_{\mathcal{F}}$, T_{σ} , and S_{σ} are the time and space complexities of checking membership in \mathcal{F} and computing σ , respectively. In contrast, we show that the problem becomes intractable for accessible set systems. We relate our results to the data mining problem of listing all support-closed patterns of a dataset and show that there is a corresponding closure operator for all datasets if and only if the set system satisfies a certain confluence property. submissions.

*A long version of this extended abstract appeared in *Theoretical Computer Science* **411**(3), 691–700, 2010.