ETH zürich

Master Thesis Proposal

Fast Poset Enumeration for Poset Function Learning

Advisor: Chris Wendler
Professor: Markus Püschel

March 18, 2021

A poset is a partially ordered set. The enumeration of all poset elements that satisfy a certain property is if major importance for various poset learning tasks. For example, in [1] the efficient enumeration of the subgraph poset is the key step for iteratively building graph classifiers by combining weak classifiers using boosting. The same idea can be also applied to graph regression.

Let $(P, \leq)$ be a poset with an unique minimal element x_{min}. The elements $x \in P$ satisfying an anti-monotone property[1] $g(x) = 1$ can be efficiently enumerated if there exists a reduction mapping $f : P \setminus \{x_{min}\} \to P$ that maps a poset element x to a smaller element $f(x) < x$ such that x *covers*[2] $f(x)$. The reduction operation defines an enumeration tree T with vertices P, edges $\{(f(x), x) : x \in P \setminus \{x_{min}\}\}$ and root x_{min}. Now, the poset can be enumerated efficiently by starting a depth first search from x_{min} in T while pruning the descendants of each x with $g(x) = 0$ that we encounter.

Reverse search [2, 3] deals with the case where the poset P is very large, e.g., P has exponential size like for $(2^N, \subseteq)$. In this case, the enumeration only works if it is possible to compute the preimage $f^{-1}(x) = \{y \in P : f(y) = x\}$ for each $x \in P$. Then, starting from x_{min} one recursively computes the preimages of $x \in f^{-1}(x_{min})$ and so on, which again results in a depth first traversal of the enumeration tree.

Your contribution The goal of this project is to implement fast poset enumeration algorithms using reverse search augmented with pruning for arbitrary posets and to evaluate them in poset function learning tasks.

In particular, your tasks are:

a) Given a poset $(P, \leq)$ represented by its cover graph (i.e., P fits into memory) and its minimal element x_{min},

[1] A property $g : X \to \{0, 1\}$ is called anti-monotone if $x \leq y$ implies $g(x) \geq g(y)$.
[2] An element $x \in P$ covers $y \in P$ if $y < x$ and there is no $z \in P$ with $y < z < x$.

1. implement a reduction mapping f and its preimage function f^{-1},
2. use f^{-1} to construct an enumeration tree T,
3. implement the efficient enumeration of T with pruning to enable poset element classification/regression,
4. experimentally evaluate your algorithm.

b) Given an arbitrary very large poset $(P, \leq)$, investigate whether it is possible to implement the efficient enumeration by reverse search [2, 3] based on a reduction mapping and its preimage function. If so

1. implement the generic reduction mapping f and its preimage function f^{-1},
2. use f^{-1} to implement the reverse search algorithm [2, 3] (plus pruning),
3. experimentally evaluate your algorithm.

Deliverables

Final report: The final report may be written in English or German. It must contain an abstract written in both English and German. It should include an introduction, an analysis of related work, and a complete documentation of all used software tools and mathematical derivations. Three copies of the final report must be delivered to the supervisor.

Reproducible experimental setup: Implementations, configuration scripts and instructions to reproduce the results reported in the thesis must be delivered in electronic form.

Presentation: The results of the thesis must be presented during a software-group seminar. The presentation is capped to 30 minutes and should give an overview as well as the most important details of the work.

Contact If you are interested in pursuing this master thesis, please contact wendlerc@ethz.ch or pueschel@ethz.ch.

References

[1] H. Saigo, S. Nowozin, T. Kadowaki, T. Kudo, and K. Tsuda. gboost: a mathematical programming approach to graph classification and regression. *Machine Learning*, 75(1):69–89, 2009.

[2] D. Avis and K. Fukuda. Reverse search for enumeration. *Discrete applied mathematics*, 65(1-3):21–46, 1996.

[3] S. Nowozin. *Learning with structured data: applications to computer vision.* PhD thesis.