Speaker: Carson Leung, UBC

Title: Exploratory Mining of Constrained Frequent Sets

Date: Wednesday April 9, 2003

Time: 1:00 p.m. - 1:50 p.m.

Room: C630

Abstract:
Data mining refers to the search for implicit, previously unknown, and potentially useful relationships or patterns (such as frequent sets) that might be embedded in data. Most of the existing data mining algorithms do not allow users to express the patterns to be mined according to their intentions, via the use of constraints. Consequently, these unconstrained mining algorithms can yield numerous patterns that are not interesting to users. Moreover, data mining is supposed to be a human-centered and exploratory process, not a one-shot exercise. In this context, we are working on a project with the overall objective of developing a practical human-centered computing environment for the efficient and effective exploratory mining of constrained frequent sets. One critical component of such an environment is the support for the dynamic mining of constrained frequent sets. Constraints enable users to impose a certain focus on the mining process. The term "dynamic" means that, in the middle of the computation, users are able to (i) change (such as tighten or relax) the constraints, and/or (ii) change the support threshold. This permits users to have a decisive influence on subsequent computations. In real-life situations, the available buffer space may be quite limited, thus adding another complication to the problem.

In this talk, we discuss how we developed the environment for such efficient and effective exploratory mining of constrained frequent sets. More specifically, in this project, we developed an algorithm, called DCF, for Dynamic Constrained Frequent-set computation. This algorithm is enhanced with a few optimizations, exploiting a light-weight structure called a segment support map. Such a structure enables DCF to (i) obtain sharper bounds on the support of patterns (or frequent sets of items), and (ii) better exploit properties of constraints. Furthermore, when handling dynamic changes to constraints, DCF relies on the concept of a delta member generating function, which exploits a special class of constraints—namely, succinct constraints—to generate precisely the sets of items that satisfy the new but not the old constraints. Our experimental results show the effectiveness of these enhancements. Lastly, to show that the exploitation of succinct constraints is not confined to DCF, we developed another algorithm that also supports the dynamic mining of constrained frequent sets, albeit in a tree-based framework.

ALL ARE WELCOME