

## EFFICIENT PROGRESSIVE SAMPLING FOR DATA MINING

Kittisak Kerdprasop, Nittaya Kerdprasop, Chaknarin Kongcharoen  
and Tippaya Tippayasot

Data Engineering and Knowledge Discovery (DEKD) Research Unit  
School of Computer Engineering  
Suranaree University of Technology  
Nakhon Ratchasima 30000, Thailand  
E-mail : kerdpras, nittaya@sut.ac.th

### Abstract

Learning from a very large data causes a performance problem even with the most efficient algorithm. Natural solution is to use a sample, but it is not obvious to determine the small but sufficient sample size. We explore two methods of progressive sampling: arithmetic and geometric. Progressive sampling is defined as a technique that starts with a small sample and repeatedly uses progressively larger samples until the performance of the resulting model does not further improve. The results of our studies show that geometric progressive sampling works well with the large data sets.

**Published in:** Proceedings of 1<sup>st</sup> KMITL International Conference on Integration of Science and Technology for Sustainable Development, Bangkok, Thailand, August 25-26, 2004. pp.313-316.