

Database Group

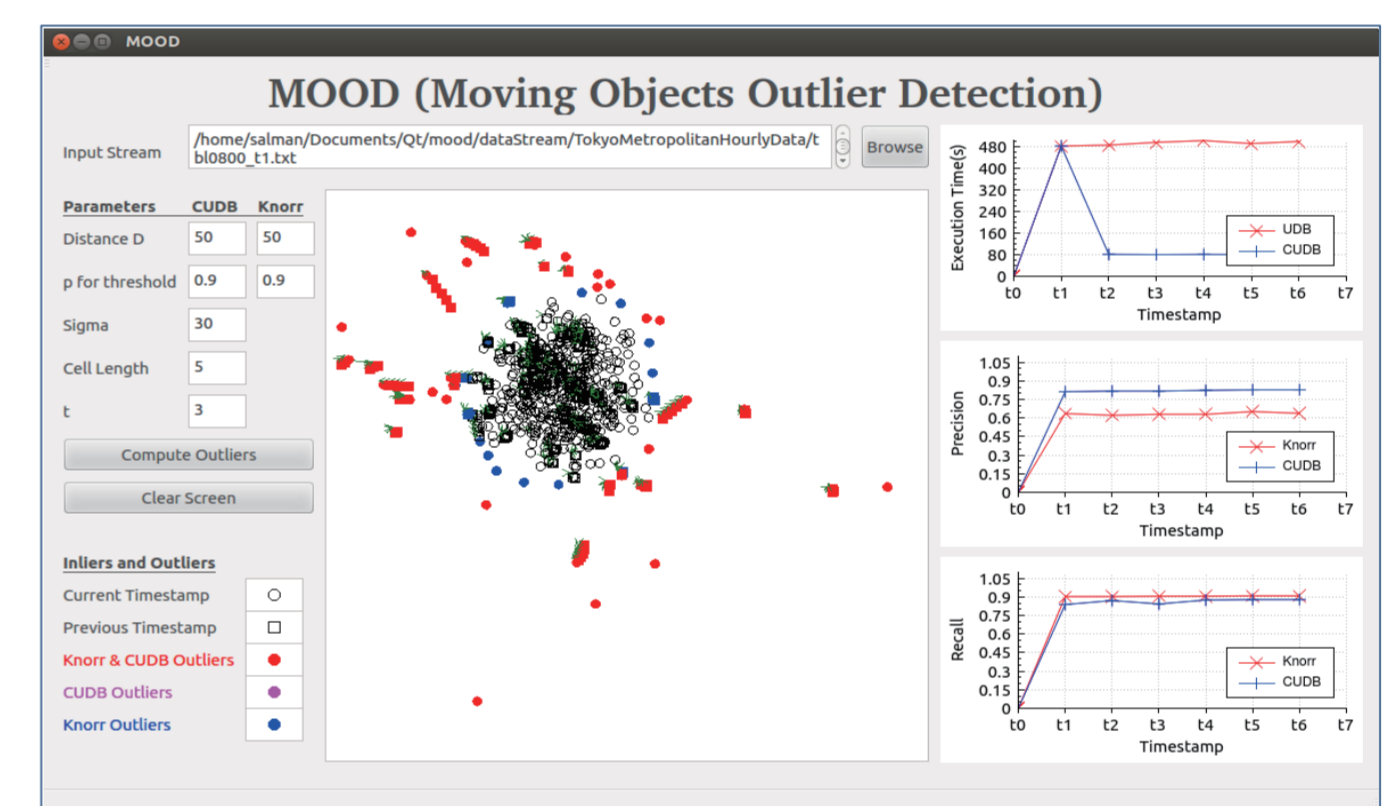
Introduction

In the field of computational sciences, management and utilization of massive data are extremely important issues. The database group in the computational informatics division is in charge of research and development in the field of data engineering. In particular, we have been engaging in the following research topics: infrastructure for integrating heterogeneous databases and various information sources, data mining and knowledge discovery technologies to discover knowledge and outliers from massive data, and scientific data management. Additionally, we also have engaged in application studies, such as development of JPV/JMA meteorological databases and knowledge discovery from the database in cooperation with Global Environment Science division and management of Lattice Data Grid in cooperation with Particle Physics division.

MOOD: Moving Objects Outlier Detection

An outlier is a data point, which is significantly different from the remaining data points in a dataset, and there are growing demands to detect outliers from various data sources to identify malicious behaviors, intrusions, etc. Besides, in many applications, data contain inherent errors due to many reasons, such as sensor errors.

MOOD is an application for continuous distance-based outlier detection from moving objects' data streams. In MOOD, it is assumed that the moving objects' data streams are uncertain, and this uncertainty is given by the Gaussian distribution. The MOOD application provides an interface which takes moving objects' data streams as input along with some parameters, and continuously produces the distance-based outliers along with graphs comparing the efficiency and accuracy of the underlying algorithms.



References:

- Salman Ahmed Shaikh, Hiroyuki Kitagawa: Continuous outlier detection on uncertain data streams. ISSNIP 2014: 1-7
- Salman Ahmed Shaikh, Hiroyuki Kitagawa: MOOD: Moving Objects Outlier Detection. APWeb 2014: 666-669

GPU Acceleration of Probabilistic Frequent Itemset Mining from Uncertain Databases

Uncertain databases have been widely developed to deal with the vast amount of data that contain uncertainty. To extract valuable information from the uncertain databases, several methods of frequent itemset mining, one of the major data mining techniques, have been proposed. However, their performance is not satisfactory because handling uncertainty incurs high processing cost. In order to address this problem, we utilize GPGPU. In this work, we propose a method for frequent itemset mining from uncertain databases using GPU. The main idea is to speed up probability computations by making the best use of GPU's high parallelism and low-latency memory.

References:

- Yusuke Kozawa, Toshiyuki Amagasa, Hiroyuki Kitagawa: GPU acceleration of probabilistic frequent itemset mining from uncertain databases. CIKM 2012: 892-901
- Yusuke Kozawa, Toshiyuki Amagasa, Hiroyuki Kitagawa: Probabilistic Frequent Itemset Mining on a GPU Cluster. IEICE Transactions 97-D(4): 779-789 (2014)

