Faster Big-data Analysis with World-class Pattern Mining Technologies

Faster Big-data Analysis with World-class Pattern Mining Technologies

- Expected to be utilized in various industries including the finance and IT sectors -

A research team led by Professor Min-Soo Kim from Department of Information and Communication Engineering developed 'GMiner' technology that can analyze big data patterns at high speed. GMiner technology exhibits performance up to 1,000 times faster than the world's current best pattern mining technology.
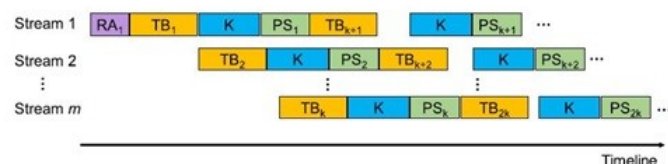

Fig 1. Multiple asynchronous GPU streams of GMiner

Pattern mining technology identifies all important patterns that appear repeatedly in the big data of various fields such as buying goods at mega-marts, banking transactions, network packets, and social networks. This technology is widely used in various industries for purposes such as determining the location of products on mega-mart shelves or recommending credit cards that match the usage patterns of consumers of different ages.

The growing importance of pattern mining has led to the development of thousands of pattern mining technologies over the past 20 years; however, due to the increasing length of big data patterns, which increased the number of analytical patterns exponentially, existing mining technologies were hindered in their analysis of data of more than ten gigabytes (GB) because they failed to complete their analysis due to insufficient computer memory or took too much time.

Traditional pattern mining technologies first found medium-length patterns and stored them in memory. When seeking a pattern that is longer than medium-length, they used a method of finding final patterns in comparison to a medium-length pattern that had been previously saved.
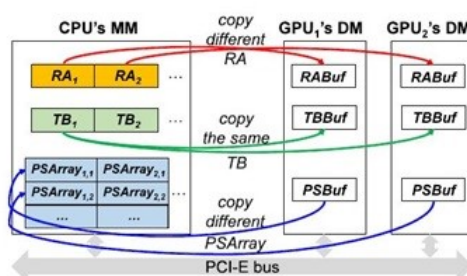

Fig 2. Data flow of GMiner using multiple GPUs

However, GMiner technology developed by the research team has succeeded in fundamentally solving the problem of existing technologies by proposing anti-intuitive techniques that combine the temporarily calculated medium-length patterns using the thousands of cores on graphics processing units (GPU) to calculate the ultimate length of patterns.

GMiner technology completely solved the chronic problem of insufficient memory suffered by conventional technologies by not storing an exponential number of medium-length patterns in memory. In addition, it solved the slow speed problem by streaming data from the main memory to the GPU while simultaneously seeking patterns using the high computational performance of the GPU.

(a) Speed-up ratio when using two GPUs (Webdocs). (b) Speed-up ratio when using two GPUs (Quest-Def).
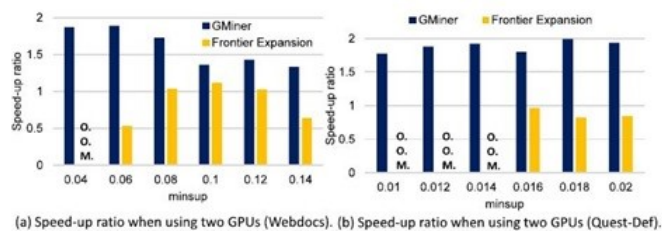
Fig 3. Speed-up ratios when using multiple GPUs for GPU-based methods

GMiner technology showed analysis performance that is a minimum of 10 times to a maximum of 1,000 times faster than conventional distributed and parallel technologies that analyzed data by using up to dozens of general home computers that have a single GPU per computer; thus, it can analyze big data on a larger scale than existing technologies. It also showed excellent expansion performance that improves performance in proportion to the number of GPUs.

Professor Kim said, "We have secured fundamental technologies that can analyze big data patterns at high speed without any problems in memory for big data accumulated in a variety of industries. By solving problems where pattern mining technologies were not properly applied to big data due to lack of memory and slow speed, this new technology can be utilized in helping companies to make efficient decisions by analyzing big data patterns in various sectors including the finance, retail, IT, and bio-related sectors."

This research outcome was published in the May 9 issue of Information Sciences, the most authoritative international journal in the field of information science.

For more information, please contact:

Professor Min-Soo Kim

Department of Information and Communication Engineering

Daegu Gyeongbuk Institute of Science and Technology (DGIST)

E-mail: mskim@dgist.ac.kr

Associated Links

Research Paper on Journal of Advanced Health Care Materials

https://doi.org/10.1016/j.ins.2018.01.046

Journal Reference

Kang-wook Chon, Min-soo Kim, et al., "GMiner: A Fast GPU-based Frequent Itemset Mining Method for Large-scale Data," Information Sciences May 2018.