

A Research on Big Data and AI Analysis Algorithm Optimization Using GPUs, Applied in Lifestyle Authentication Application

TRAN VAN SANG¹ Ryosuke KOBAYASHI² RIE Shigetomi YAMAGUCHI³
Toshiyuki NAKATA⁴

1. Background

With the significant increase of computer performance, in recent years, many complex human-like tasks have been resolved by computer software in reasonable time. These tasks include visual object recognition, speech to text interpretation, human face authentication, etc. . . . However, computer performance is going to reach the limit as the CMOS transistor size nears the limit. On the other hand, the amount of data which need to be processed is incredibly growing up under Internet of Thing, Industrialization 4.0, social network era [2], which leads to the demand of higher scalability on current Big Data, AI analysis algorithms.

2. Our research

Our research investigated on finding scaling solution for Big Data, AI analysis problems. The whole development is composed of 2 phases: acceleration by GPU and distributed computing application. This research focuses on the former topic with the integration in an specific application named Lifestyle Authentication. Lifestyle Authentication [3] is an authentication method taking advantage of user's lifestyle information to improve system authenticity. A real-world dataset from MangaOne application was used in this research to achieve more real-life optimization and model evaluation result. The data includes 41,638,144 log records of 49,261 users over 8,502 chapters of 79 series in 9 months from 2014 December 18 to 2015 September 10.

3. Methodology, result and conclusion

The research was conducted in 3 steps: raw data manipulation, model choosing with evaluation, and speed optimization using GPU. Firstly, data was aggregated and filtered in data manipulation phase to make it possible to be applied in AI model while keeping sufficient information. To apply some supervised learning models, fake data, i.e negative samples, were also generated in such a way that generates data more likely to the original dataset. Next, many models were tried being fitted with input data set. Graph based approach, Multilevel hypergraph partition, was one of the

most promising approach, but it was too complicated to be fully investigated within this research scope. Finally decision tree and binary regression were chosen with the highest correctness at 85%, False Rejection Rate and False Acceptance Rate be around 10%. These figures were 1.14 times and 1.3 times better than previous result of R. Kobayashi [1]. After that, speed optimization phase was developed to improve binary regression training speed using GPU acceleration. The first step of the optimization is to speed up the core calculation with QR decomposition in training phase using GPU, while keeping other wrapping portion being processed in CPU. In this step of the optimization, memory movement between CPU and GPU was replicated multiple times and took relatively high timing portion at more than 20% of the whole training time. As a result, our current optimization implementation speed up the training phase 1.27 times regarding the fact that this step theoretically speeds up at most 2.2 times due to the Amdahl's law. . On the other hand, we also found out that input matrix shape strongly affects speed up rate of the core calculation via an experiment which shows that the optimization speeds up 80 times with square-shaped input matrix while it speeds up only 1.65 times for input matrix with the same shape as in our application. For these reasons, it is believed that there is still more space for the speed optimization in AI analysis and Big Data, to be developed. We strongly believe that they will be applicable in our following researches.

References

- [1] RYOSUKE KOBAYASHI and RIE Shigetomi Yamaguchi. A utilization of manga browsing and other usage history in manga application for an authentication method. In *A Utilization of Manga Browsing and Other Usage History in Manga Application for an Authentication Method*, volume Symposium on Cryptography and Information Security of SCIS 2017, Naha, Japan, 1 2017. The Institute of Electronics, Information and Communication Engineers, The Institute of Electronics, Information and Communication Engineers.
- [2] Steve Lohr. The age of big data. *New York Times*, 11(2012), 2012.
- [3] 小林 良輔, 疋田 敏朗, 鈴木 宏哉, and 山口 利恵. 行動センシングログを元にしたライフスタイル認証の提案. In コンピュータセキュリティシンポジウム 2016 論文集, number 2 in 1, pages 1284–1290, oct 2016.