www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issues 6 June 2016, Page No. 17107-17109

Survey :Various Prediction Algorithms for Chemical Bond Formation on GPU

Manjiri K. Kulkarni, Dr. J. S. Umale

Department of Computer Engineering
Pimpri Chinchwad College of Engineering
Pune-44
manjirik2310@gmail.com
Professor, Department of computer Engineering,
Pimpri Chinchwad Collage of Engineering
Pune-44

jayantumale@gmail.com

ABSTRACT
Nowadays, the persuit of prediction of chemical bond formation among researchers is evident to discover new drugs or for new discovery of chemicals. Predictive modeling gives statistics to predict outcomes. Most of time the event one wants to predict is in the future, but predictive modeling applicable to any type of event, regardless of when it occurred. In this paper, we focused on comparative analysis of various Prediction algorithms to estimate the best algorithm for the prediction of chemical bond formation with observations. The performances of these techniques are compared and it is observed that parallel Genetic Algorithm provides better performance results in accuracy and speedup as compared to other prediction techniques on GPU.

Keywords
GPU Computing, Chemical Bonds, Prediction Algorithms.

INTRODUCTION
Predictive analytics is proving itself powerful because advanced algorithms can take a near-unlimited number of factors into account, provide deep insights into variation, and scale to meet the needs of even the largest company. Current methods of data analysis which work based on reviewing the statistical graphs, has many limitations to predict the performance and availability of the produced parts. So in current world, analysts are attending to use superior patterns to increase availability. Recent years, data mining is considered as one of the common methods for processing and discovering the hidden patterns [1].

Nowadays, Graphic Processing Units (GPU) have drawn increasing popularity for high performance computing. The NVIDIA Compute Unified Device Architecture (CUDA) summarizes GPU as a general-purpose multithreaded SIMD (single instruction, multiple data) architectural model, and provides a C-like interface supported by a compiler and a runtime system for GPU programming [5]. As in the case of CPU programming, ensuring that a GPU application efficiently utilises computational resources is a cardinal goal. Following strategies are useful for optimization of GPU applications :

1. Maximising parallel execution
2. Optinising memory usage to achieve maximum memory bandwidth
3. Optimising usage of instructions to achieve maximum instruction throughput

Performance prediction of computer systems plays an important role in many computer science fields:

While parallel performance it is important to consider overhead introduced by communication, thread spawning, and synchronization. Performance loss will be there, if the overhead is high.

Thus, an important question in this process is to evaluate whether the optimization brings any performance improvements. The answer is usually computed using a performance model which is an abstraction of the target hardware.

Prediction algorithm :
Prediction means guessing future trends from past information. The purpose of a prediction algorithm is to forecast future values based on our present records. [3] Some common tools for prediction include: neural networks, regression, Support Vector Machine (SVM), and discriminant analysis. Recently, data mining techniques such as neural networks, fuzzy logic systems, genetic algorithms and rough set theory are used to predict control and failure detection tasks[4]. In this paper, the algorithms will forecast a probability for the given data situation. If the probability is equal to 1 it means the data (part) is normal, otherwise if the probability is equal to 0 the data (part) is considered non-conventional.

LITERATURE SURVEY

Following Table shows review on previous works--

Sr. No.	Algorithms	Advantages	Disadvantages
1	Support Vector Machine	Based on sound mathematics theory Learning result is	Difficult to understand the learned function Not easy to incorporate domain

		more robust	knowledge
2	Decision Tree	Require relatively little effort from users for data preparation Easy to interpret	Complex and time-consuming Possibility of invalid relationships
3	Artificial Neural Network	very well on pattern recognition tasks with a large amount of training data.	Greater computational burden and empirical nature of model development.
4	Parallel Genetic Algorithm	Intrinsically parallel Successfully finding optimal or very good results in a short period of time Global Optimum	Fitness function must be carefully considered. If population size is too small, it does not give good solution.

Machine learning methods are being used by several researchers for successfully predicting new drug. Various algorithms are discussed as follows :

1. Support Vector Machine (SVM)

The support vector machine (SVM) is a supervised learning method that generates input-output mapping functions from a set of labeled training data. Support Vector Regression ignores any training data that is sufficiently close to the model prediction. With the growth of massive data, the computational complexity of running SVM algorithms also increases drastically, which may seriously limit the practical usage of SVM in large-scale data based applications. In order to avoid this, both efficient optimization algorithms and powerful computing are required. For this it can be implement on Graphics Processing Units (GPUs). But it is Difficult to understand the learned function or mapping function[7].

2. Decision Tree

Decision tree falls under supervised learning techniques as we have known labels in the training data set in order to train the classifier[3]. The Traditional Algorithm for learning decision trees is implemented using information gain as well as using gain ratio. For large number of parameters values, it will take more time to compute. This difficulty can be address by Graphics processing Units (GPUs), so that information gain and gain ration can be calculated parallely. But there may be possibility that dependency between parameters leads to decrease in parallelization.

3. Artificial Neural Network

For large amount of data requires efficient data processing methods[1]. There are most commonly used methods for data processing applications like Artificial neural networks (ANNs) [2]. Artificial neural networks are computational models inspired by the nervous systems in nature and have found extensive utilization in solving many complex real-world problems [6].

4. Parallel Genetic Algorithm

Genetic Algorithms (GAs) are adaptive heuristics search algorithm based on the evolutionary ideas of natural selection ad genetics. The fitness function is defined such that it takes an individual as its input and gives its fitness value. It is used to find the optimal combination of considered parameters so that prediction of future things will be accurate[4]. To increase the performance of GA, crossover and mutation steps are applied in parallel. It is referred as Parallel Genetic Algorithm. In the prediction of chemical bond formation, Parallel GA will be used to find the optimal combination of considered parameters so that those combinations are the accurate values for chemical bond formation. As the chemical dataset increases, it may take lifetime of researchers. To address this challenge genetic algorithm can be mapped to CUDA programming on GPUs so that performance and accuracy will get increased[5].

Observations:

From previous paper work[7], it is observed that Support Vector Machine learning can be accelerated by utilizing the parallel computing power of GPU. But the difficulty is in mapping function as SVM based on mathematical theory so all the theory of work should get map with mapping function. So for optimization as well as prediction of chemical bond formation, SVM keeps drawbacks.

Although Decision Tree classifying the data is that they are simple to understand and interpret but there are some disadvantages such as, Most of the algorithms (like ID3 and C4.5) require that the target attribute will have only discrete values and as decision trees use the "divide and conquer" method, they tend to perform well if a few highly relevant attributes exist, but less so if many complex interactions are present [3]. By studying decision tree, it is observed that decision tree will not suitable to predict chemical bond formation.

From the privous paper work[1,2], in the ANN parallelism is achieved within the calculations from hidden layer to output layer, Based on that, they propose an efficient GPU implementation of large scale recurrent neural networks with a fine-grained two-stage pipeline architecture.

From the previous paper work [4,5], it is observed that Parallel Genetic Algorithm can be easily map on GPU than other prediction algorithms. Time required to find fitness function is very less than to find mapping function in SVM. Genetic Algorithm finds global optima and well working with large dataset. So prediction of chemical bond formation as well as optimise combination of values of parameters will be possible.

CONCLUSION

From this survey work, it is clear that parallel genetic algorithm is more suitable for chemical application to the prediction of chemical bond formation.

REFERENCES

[1] I. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," Journal of microbiological methods, vol. 43, no. 1, pp. 3–31, 2000.

[2] R. Men´endez de Llano and J. L. Bosque, "Study of neural net training methods in parallel and distributed architectures," Future Generation Computer Systems, vol. 26, no. 2, pp. 267–275, 2010.

[3] Ahmad Ashari, P.Palazzari, Iman Paryudi, A Min Tjoa, Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 11, 2013.

[4] Fauzi Mohd Johar, Farah Ayuni Azmin, Mohamad Kadim Suaidi, A Review of Genetic Algorithms and Parallel Genetic Algorithms on Graphics Processing Unit (GPU), 2013 IEEE International Conference on Control System, Computing and Engineering, 29 Nov. - 1 Dec. 2013, Penang, Malaysia, 978-1-4799-1508-8/13/$31.00.

[5] Archit Somani and Dhirendra Pratap Singh, Parallel Genetic Algorithm for solving Job-Shop Scheduling Problem Using Topological Sort, 978-1-4799-6393-5/14/$31.00 ©2014 IEEE.

[6] Boxun Li, Erjin Zhou, Bo Huang, Jiayi Duan, Yu Wang, Ningyi Xu, Jiaxing Zhang, Huazhong Yang, Large Scale Recurrent Neural Network on GPU, 2014 International Joint Conference on Neural Networks (IJCNN) July 6-11, 2014, Beijing, China.

[7] Tianyao Sun, Hanli Wang, Yun Shen, Jun Wu, "Accelerating Support Vector Machine Learning with GPU-based MapReduce", 978-1-4799-8697-2/15 $31.00 © 2015 IEEE DOI 10.1109/SMC.2015.161.