An Unlimited Vocabulary Korean Voice Interface by Using Grid Engine-based Training Speed Improvement

Donghyun Lee, Kwang-Ho Kim, Hee-Eun Kang, Shang-Ho Wang, Sung-Yong Park, Ji-Hwan Kim[1]

Abstract This demonstration highlights the improvements on training speed when the Grid Engine is applied to the development of an unlimited Korean vocabulary voice interface. This paper proposes a distributed architecture using the open source software, Sun Grid Engine (SGE). Testing has shown that the training speed is 5 times faster using this proposed implementation while still able to maintain a similar speech recognition rate.

1. Overview

This paper proposes a distributed architecture using the open source software, Sun Grid Engine (SGE) [1]. SGE is a batch scheduling software which distributes jobs submitted by users to cluster nodes. The main components of SGE are as follows: a master host, submit hosts, and execution hosts (slave). Users submit jobs to the queue via the submit host. The master host selects one of the execution hosts which has the least load and assign jobs to it. The number of jobs that can be concurrently processed by SGE is equal to the sum of the execution hosts' cores.

The Korean speech corpus for training was from ETRI and SiTEC corpus (approximately 160,000 speech files). The evaluation corpus was collected by two users. Testing was conducted regarding the Korean acoustic model based on Deep Neural Network using SGE installed in Amazon EC2 [2]. In order to compare the training speed, a single server (the baseline) was used to conduct the same amount of training. In this configuration, the single server has the same specifications as the master server in Table 1.1.

[1] Donghyun Lee, Kwang-Ho Kim, Hee-Eun Kang, Shang-Ho Wang, Sung-Yong Park, Ji-Hwan Kim
Dept. of Computer Science and Engineering, Sogang University, e-mail: {redizard , kimkwangho, heun831, sangho362, parksy, kimjihwan}@sogang.ac.kr

Table 1.1 Specifications of master and slave servers implemented in Amazon EC2 servers

Host type	No. of servers	HW specification	Server type
Master	1	No. of core: 32 Memory: 244GB No. of GPU: 0	r3.8xlarge
Slave	5	No. of core: 8 Memory: 15GB No. of GPU: 1	g2.2xlarge

Comparison results for the various system configurations are shown in Table 1.2 in terms of word recognition rate and processing time. These results indicate that the proposed system configuration (multi-serves based on GE) successfully increased training speed five times while maintaining word recognition rate.

Table 1.2 Comparison of the word recognition rate and processing time according to the system configuration

System configuration	Word recognition rate (%)		Processing time (hour)
	Man	Woman	
Single server	58.66	56.74	41
Multi servers based on GE (proposed)	58.75	56.78	8

2. Conclusion

This demonstration suggested the improvements on training speed when the Grid Engine is applied to the development of an unlimited Korean vocabulary voice interface. This architecture is implemented through 6 severs on the Amazon Elastic Compute Cloud (Amazon EC2): one Master server and five Slave servers. The Master server utilizes 32 CPU cores and 244GB memory, and each of the five Slave servers uses 8 CPU cores, 15GB memory, and one GPU. The deep neural network-based acoustic models are trained by a 320-hour Korean speech corpus. Testing has shown that the training speed is 5 times faster using this proposed implementation while still able to maintain a similar speech recognition rate. In this demonstration, the results of our proposed training speed improvement implementation are introduced via an unlimited Korean vocabulary voice interface.

References

1. Gentzsch, W. (2001) Sun Grid Engine: Towards Creating a Compute Power Grid. *Proc. the first IEEE/ACM International Symposium on Cluster Computing and the Grid*, Brisbane, Australia, pp. 35-36 .

2. Juve, G., Deelman, E., Berriman, G.B., Berman, B.P., Maechling, P. (2012) An evaluation of the cost and performance of scientific workflows on Amazon EC2. *Journal of Grid Computing*, vol 10, no 1, pp. 5-21.