

Implementation of a large-scale language model adaptation in a cloud environment

Kwang-Ho Kim · Dae-Young Jung · Donghyun Lee · Hyuk-Jun Lee ·
Sung-Yong Park · Myoung-Wan Koo · Ji-Hwan Kim ·
Jeong-sik Park · Hyung-Bae Jeon · Yun-Keun Lee

© Springer Science+Business Media New York 2013

Abstract This paper presents a system of large-scale language model adaptation for daily generated big-size text corpus using MapReduce in a cloud environment. Our large-scale trigram language model, consisting of 800 million trigram counts, was successfully implemented by a new approach using a representative cloud service (Amazon EC2), and a representative

This paper is an extended version of the 3rd International Conference on Intelligent Robotics, Automations, Telecommunication Facilities, and Applications (IRoA 2013).

K.-H. Kim · D.-Y. Jung · D. Lee · H.-J. Lee · S.-Y. Park · M.-W. Koo · J.-H. Kim (✉)
Department of Computer Science and Engineering, Sogang University, Seoul, Republic of Korea
e-mail: kimjihwan@sogang.ac.kr

K.-H. Kim
e-mail: kimkwangho@sogang.ac.kr

D.-Y. Jung
e-mail: dyjung@sogang.ac.kr

D. Lee
e-mail: redizard@sogang.ac.kr

H.-J. Lee
e-mail: hyukjunl@sogang.ac.kr

S.-Y. Park
e-mail: parksy@sogang.ac.kr

M.-W. Koo
e-mail: mwkoo@sogang.ac.kr

J.-s. Park
Department of Intelligent Robot Engineering, Mokwon University, Daejeon, Republic of Korea
e-mail: parkjs@mokwon.ac.kr

H.-B. Jeon · Y.-K. Lee
Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

H.-B. Jeon
e-mail: hbjeon@etri.re.kr

Y.-K. Lee
e-mail: yklee@etri.re.kr