Database Group

Scalable Graph Analysis over Intel Xeon Phi Coprocessors

The structural graph clustering method *SCAN* is successfully used in many applications since it detects not only densely connected nodes as clusters but also extracts sparsely connected nodes as hubs or outliers (Fig. 1). However, it is difficult to apply SCAN to large-scale graphs since SCAN needs to evaluate the density for all adjacent nodes included in the graph. In this work, so as to address the above problem, we present a novel algorithm *SCAN-XP* that performs on Intel Xeon Phi coprocessors. We designed SCAN-XP to make the best use of many cores in the Intel Xeon Phi by employing the following approaches: First, SCAN-XP avoids the bottlenecks that arise from parallel graph computations by providing good load balances among the cores. Second, SCAN-XP effectively exploits 512 bit SIMD instructions implemented in each core to speed up the density evaluations. As a result, SCAN-XP runs approximately 100 times faster than SCAN; for the graphs with 100 million edges, SCAN-XP is able to perform in a few seconds (Fig. 2).

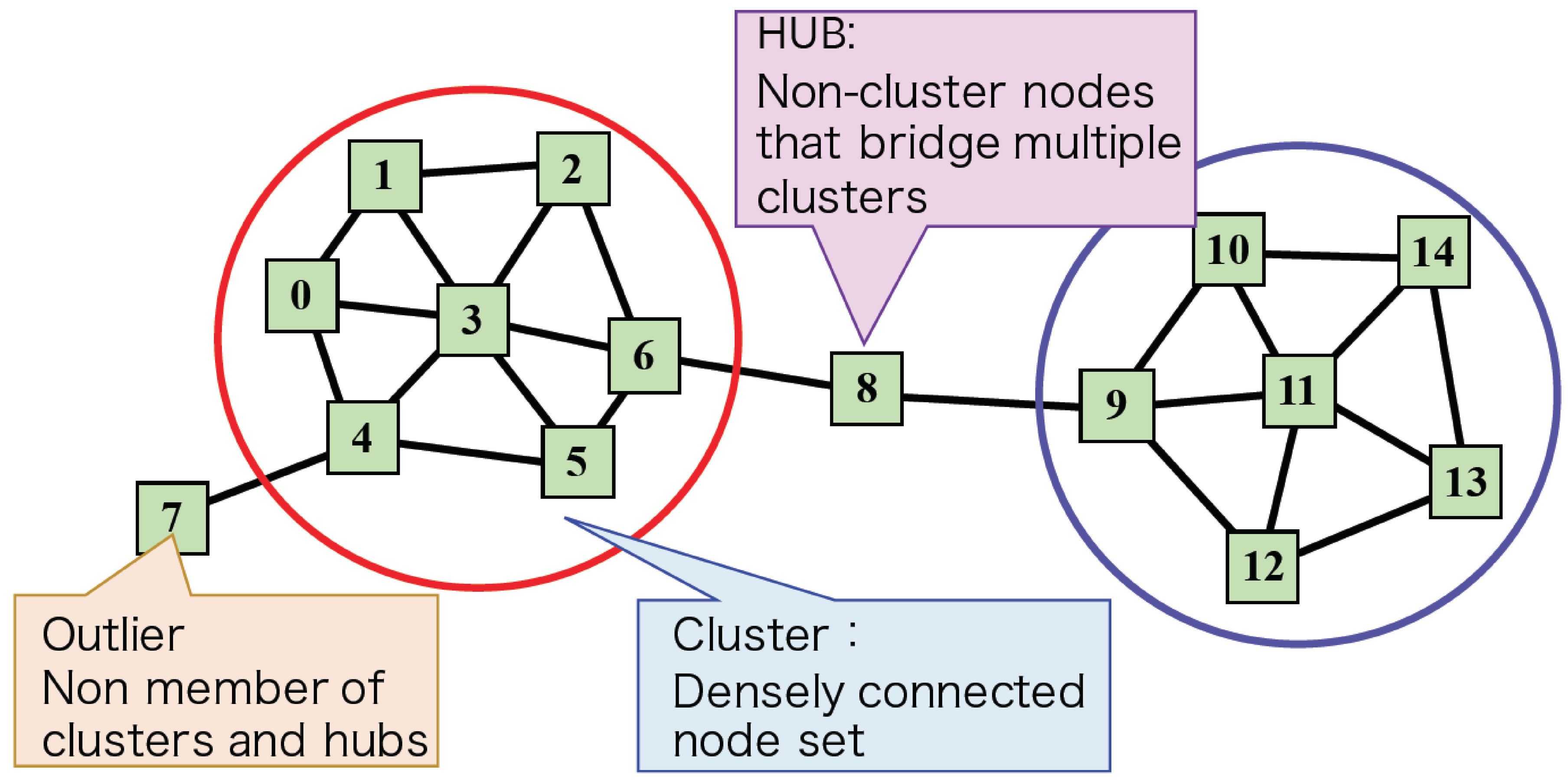

Fig. 1: Structural Graph Clustering SCAN

Table. 1: Real-world Dataset

Dataset	# of nodes	# of edges
com-youtube	1,134,890	2,987,624
web-BerkStan	685,230	6,649,470
soc-Pokec	1,632,803	22,301,964
com-LiveJournal	3,997,962	34,681,189
soc-LiveJournal1	4,846,609	42,851,237
com-Orkut	3,072,441	117,185,083
webbase2001	115,554,441	854,809,761

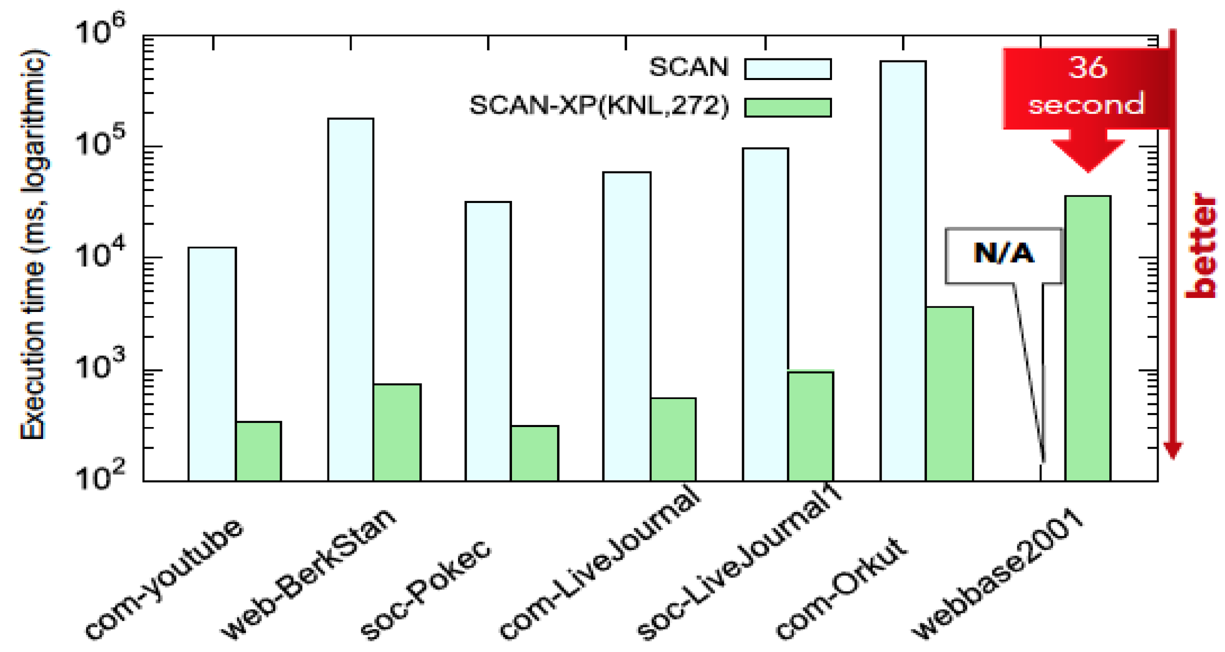

Fig. 2: Overall performances

Noise-robust sleep stage scoring for mice using deep learning & big data

Sleep stage scoring for mice is one of the most basic analyses in sleep research; however, this analysis is time-consuming and requires considerable expertise and effort. Although several studies have proposed automated scoring methods, they do not achieve robustness against noise in biological signals enough for research uses. To develop a noise-robust scoring method, we employ the following approaches.

1) Employing convolutional neural networks (CNN) & long short-term memory (LSTM), which can locate the feature of both biological signals and noise in them.
2) Training the model using noisy biological signals obtained from over 3000 mice.

Thank to these improvements, the proposed method achieved scoring accuracy of more than 95% for noisy biological signals. This result indicates that our method is practical enough for sleep research uses.

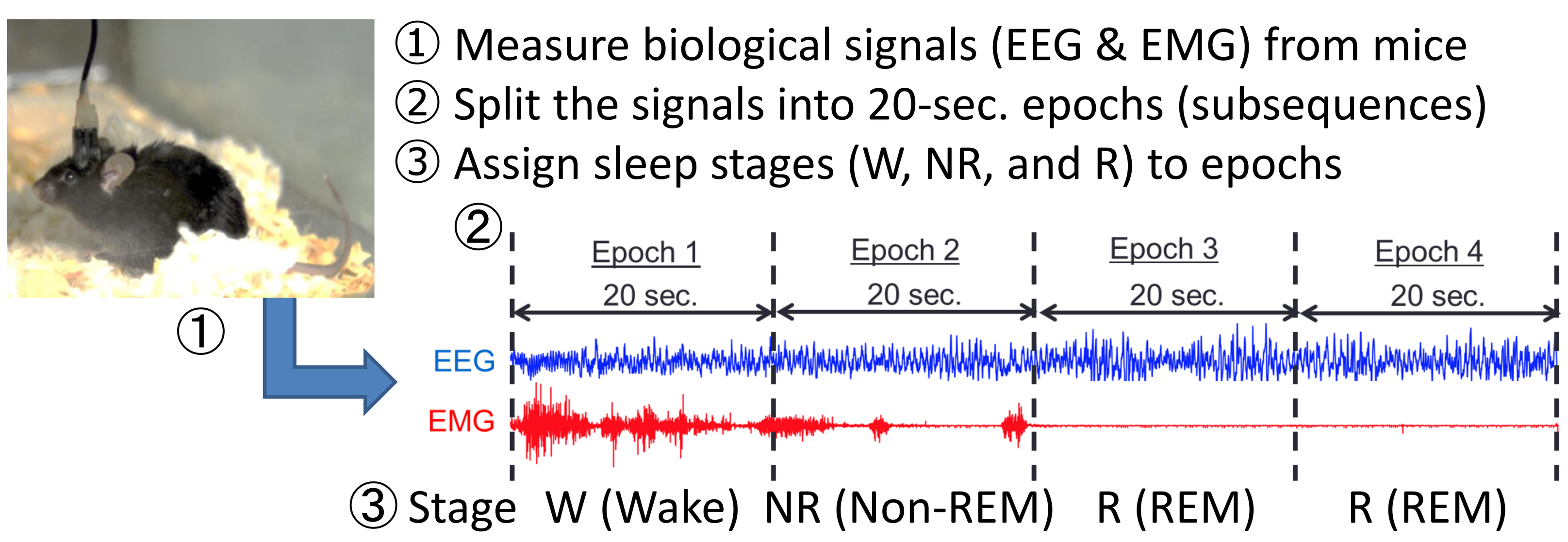

Fig. 3: Procedure of sleep stage scoring

Table 2: Feature of each stage

Stage	Peak Freq. of EEG	Amplitude of EMG
W	7-11 Hz	Large
NR	1-6 Hz	Small
R	7-11 Hz	Smallest

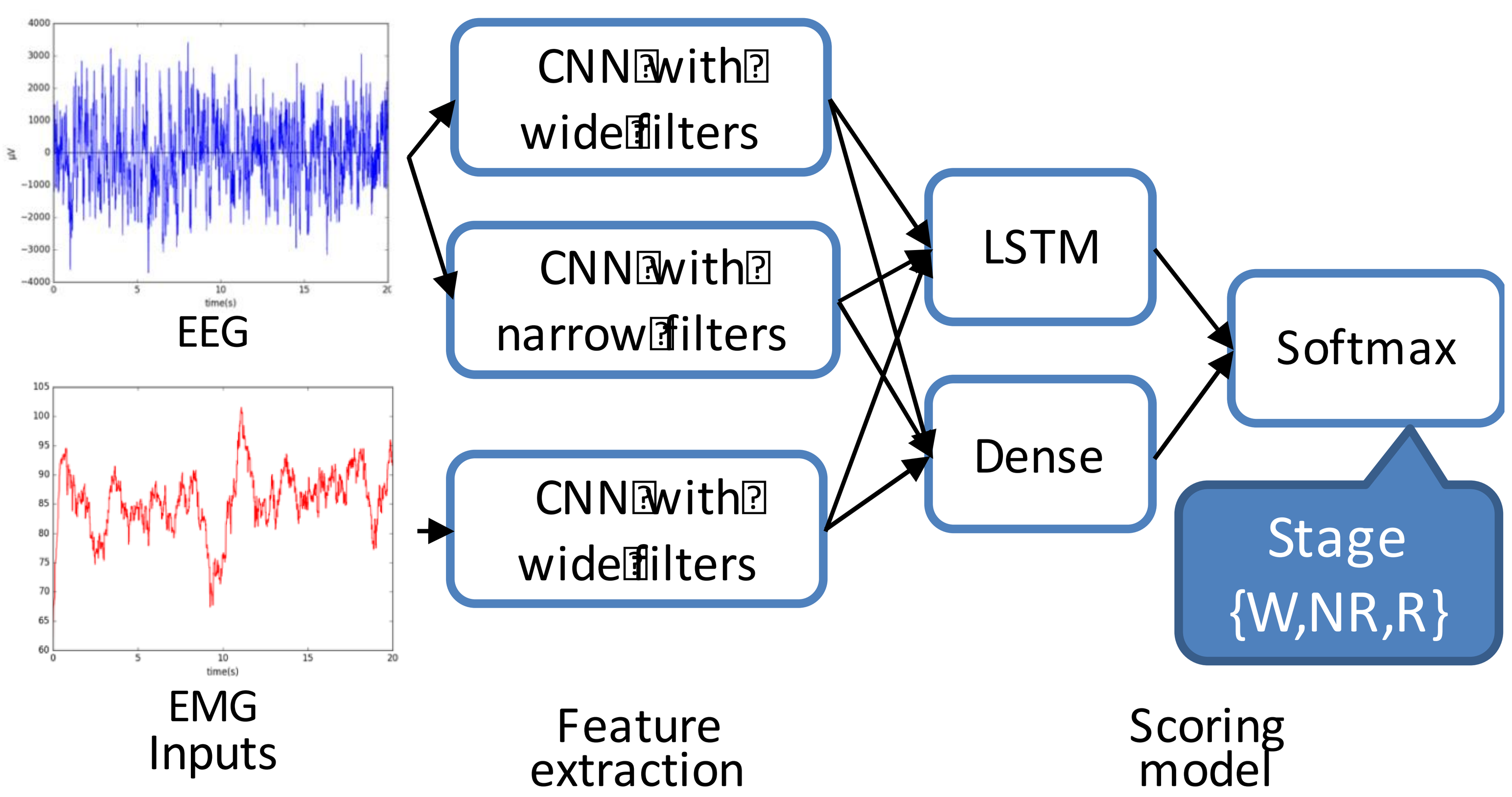

Fig. 4: Structure of the proposed system

contact address: {kitagawa, amagasa, shiokawa, horie}@cs.tsukuba.ac.jp