

INTISARI

PARALELISASI NAÏVE BAYES CLASSIFIERS UNTUK ANALISIS SENTIMEN TWEET MENGGUNAKAN GRAPHICS PROCESSING UNIT (GPU)

ENY MARIA
12/340284/PPA/04069

Penggunaan komputer untuk menyelesaikan masalah telah dilakukan untuk segala bidang pekerjaan. Seiring dengan hal tersebut, dituntut proses komputasi yang semakin cepat. Paralelisasi menjadi sebuah pilihan setelah pemrosesan sekuensial mengalami berbagai masalah dan keterbatasan. Metode paralelisasi dalam penelitian ini menggunakan paralelisasi pada GPU, untuk melakukan analisis sentimen data yang diperoleh dari internet. Data diambil dari *micro-blogging* yang pada saat ini menjadi alat komunikasi paling populer dan digemari oleh para pengguna internet.

Metode yang digunakan untuk membangun model klasifikasi atas data training dalam penelitian ini adalah metode *Naïve Bayes Classifiers*. Data training dikumpulkan dengan memanfaatkan fasilitas *crontab* dengan *query emoticon* dan akun media nasional yang terhubung ke Twitter API. Data yang terkumpul akan melewati *preprocessing* sebelum dilakukan *training*. Fitur pembobotan yang digunakan adalah *term frequency* dengan TF-IDF. Semua data yang digunakan dalam penelitian ini adalah tweet yang disampaikan dalam Bahasa Indonesia. Model paralelisasi yang dilakukan dengan menggunakan komputer *single instruction multiple data* (SIMD). Proses paralelisasi yang dilakukan menggunakan Nvidia GeForce 930M. Proses *pre-processing*, *training*, dan pencarian nilai probabilitas *term* dilakukan secara sekuensial. Proses paralel yang dilakukan untuk menghitung nilai V_{MAP} dengan perintah *multiply* atau mengalikan nilai probabilitas *term* setiap *tweet* pada setiap kategori yang dilakukan pada *thread*. Hasil implementasi diperoleh akurasi 96,61% untuk klasifikasi secara sekuensial maupun paralel. Proses klasifikasi yang dilakukan secara paralel mencapai 11 kali lebih cepat jika dibandingkan dengan proses klasifikasi secara sekuensial pada percobaan dengan 1000 tweet.

Kata kunci - - Paralelisasi, GPU, Analisis Sentimen, Micro-blogging.

ABSTRACT

PARALLELIZATION OF NAÏVE BAYES CLASSIFIERS USING GRAPHICS PROCESSING UNIT (GPU) FOR SENTIMENT ANALYSIS TWEET

ENY MARIA
12/340284/PPA/04069

Computers be the one of tools to solve the problem and have been made to all areas of work. Along with this, the required computation process accelerated. Parallelization became an option after experiencing sequential processing problems and limitations. Parallelization method in this research uses the GPU parallelization, to perform sentiment analysis of data obtained from the internet. Data taken from micro-blogging which is currently the most popular means of communication and favored by users.

The method used to build the model of the classification of the training data in this research is Naïve Bayes Classifiers. Training data collected by utilizing the crontab by querying emoticons and national media accounts are linked to the Twitter API. The data collected will pass through certain preprocessing prior to training. Feature weighting used is the term frequency with TF-IDF. All data used in this study are presented in the Indonesian tweet. Single instruction multiple data (SIMD) is the architecture for this parallelization. Parallelization process is performed using the GeForce 930M. Pre-processing, training, and search the term probability use sequential. Parallelization for count V_{MAP} value with code multiply the term probability each tweet for all class that do in the thread. The results obtained by the implementation of the classification accuracy of 96.61% on a sequential or parallel manner. The time of the classification process in parallel 11 times faster than classification process sequentially on trial for 1000 tweet.

Keywords: Parallelization, GPU, Sentiment Analysis, Microblogging