

ABSTRACTS PARAMO

HIGH PERFORMANCE QUERIES USING COMPRESSED BITMAP INDEXES

Beytullah Yildiz

Data that often contain unchanging records is becoming increasingly important. Many data sources, such as historical archives, sensor readings, health systems, and machine logs, do not change frequently but are constantly increasing. For this reason, the need to process such datasets more quickly has emerged. Bitmap index that can benefit from multicore and multiprocessor systems is designed to process data that has grown over time but does not change frequently. It has a well-known advantage, particularly in low cardinality data queries. Data such as gender, age, marital status, postal code and even date with low cardinality occupy an important place in datasets. Furthermore, the bitmap index using the compression algorithm can be applied efficiently even if the data has a high cardinality. In this study, bitmap index is introduced to improve queries using optimal encoding and it has been shown to perform up to 20x faster queries with an appropriate encoding for data containing frequently unchanging records in a performance comparison against a commonly used relational database system.

LEARNING QUALITY IMPROVED WORD EMBEDDING WITH ASSESSMENT OF HYPERPARAMETERS

Beytullah Yildiz and Murat Tezgider

Deep learning practices have a large impact on many areas. Big data and key hardware developments in GPU and TPU are the main reasons behind deep learning success. The recent progress in the text analysis and classification using deep learning has been significant as well. The quality of word representation that has become much better by using methods such as Word2Vec, FastText and Glove has been important in this improvement. In this study, we aimed to improve Word2Vec word representation, which is also called embedding, by tuning its hyperparameters. The minimum word count, vector size, window size, and the number of iterations were used to improve word embeddings. We introduced two approaches, which are faster than grid search and random search, to set the hyperparameters. The word embeddings were created using documents with approximately 300 million words. A deep learning classification model that uses documents consisting of 10 different classes was applied to evaluate the quality of word embeddings. A 9% increase in classification success was achieved only by improving hyperparameters.

HUGO: A CLUSTER SCHEDULER THAT EFFICIENTLY LEARNS TO SELECT COMPLEMENTARY DATA-PARALLEL JOBS

Lauritz Thamsen, Ilya Verbitskiy, Sasho Nedelkoski, Vinh Thuy Tran, Vinicius Meyer, Miguel G. Xavier, Odej Kao, and César A. F. De Rose

Distributed data processing systems like MapReduce, Spark, and Flink are popular tools for analysis of large datasets with cluster resources. Yet, users often overprovision resources for their data processing jobs, while the resource usage of these jobs also typically fluctuates considerably. Therefore, multiple jobs usually get scheduled onto the same shared resources to increase the resource utilization and throughput of clusters. However, job runtimes and the utilization of shared resources can vary significantly depending on the specific combinations of co-located jobs.

This paper presents Hugo, a cluster scheduler that continuously learns how efficiently jobs share resources, considering metrics for the resource utilization and interference among co-located jobs. The scheduler combines offline grouping of jobs with online reinforcement learning to provide a scheduling mechanism that efficiently generalizes from specific monitored job combinations yet also adapts to changes in workloads. Our evaluation of a prototype shows that the approach can reduce the runtimes of exemplary Spark jobs on a YARN cluster by up to 12.5%, while resource utilization is increased and waiting times can be bounded.

FLY: A DOMAIN-SPECIFIC LANGUAGE FOR SCIENTIFIC COMPUTING ON FAAS

G. Cordasco, M. D'Auria, A. Negro, V. Scarano, and C. Spagnuolo

Cloud Computing is widely recognized as distributed computing paradigm for the next generation of dynamically scalable applications. Recently a novel service model, called Function-as-a-Service (FaaS), has been proposed, that enables users to exploit the computational power of cloud infrastructures, without the need to configure and manage complex computations systems. FaaS paradigm represents an opportunity to easily develop and execute extreme-scale applications as it allows fine-grain decomposition of the application with a much more efficient scheduling on cloud provider infrastructure.

We introduce FLY, a domain-specific language for designing, deploying and executing scientific computing applications by exploiting the FaaS service model on different cloud infrastructures. In this paper, we present the design and the language definition of FLY on several computing (local and FaaS) back-ends: Symmetric multiprocessing (SMP), Amazon AWS Lambda, Microsoft Azure Functions, Google Cloud Functions, and IBM Bluemix/Apache OpenWhisk. We also present the first FLY source-to-source compiler, publicly available on GitHub, which supports SMP and AWS back-ends.



Submission deadline for the Euro-Pas PhD Symposium has been extended to 27 May, 2023. Click here for more information - <https://t.co/wWxisiCJSC>

17.05.2023 - 11:51

The Euro-Par PhD Symposium is a welcoming and supportive forum for PhD students to present their work. Click here for more information: <https://t.co/wWxisiCJSC>

04.04.2023 - 09:25

Submit your paper for EURO-PAR 2023 Workshops and Minisymposia! [Click here for more information.](https://t.co/UEseXWb3Dz) <https://t.co/UEseXWb3Dz>

07.03.2023 - 08:18

Abstract s
24 Feb, 20
<https://t.cc>



CONTACT US

HOSTS



SPONSORS



SHARE ON:



