

DATA PARTITIONING VIA HIGH DIMENSIONAL MODEL REPRESENTATION BY USING PARALEL COMPUTING

SUMMARY

If the values of a multivariate function $f(x_1, x_2, \dots, x_N)$ are given at only a finite number of points in the space of its arguments and an interpolation which employs continuous functions is considered standard multivariate routines may become cumbersome as the dimensionality grows. This urges us to develop a divide-and-conquer algorithm which approximates the function. The given multivariate data is partitioned into low-variate data. This approach is called High Dimensional Model Representation (HDMR). However the method in its current form is not applicable to problems having huge volumes of data. With the increasing dimension number and the number of the corresponding nodes, the volume of data in question reaches such a high level that it is beyond the capacity of any individual PC because huge volume of data requires much higher RAM capacity. Another aspect is that the structure of equalities used in the calculation of HDMR terms varies according to the dimension number of the problem. The number of loops in the algorithm increases with the increasing dimension number. In this work, as a first step, the equations used are modified in such a way that their structure does not depend on the dimension number. With the newly obtained equalities, the method becomes appropriate for parallelization. Due to the parallelization, the RAM problem arising from problems with high volume of data is solved. Finally, the performance of the parallelized method is analyzed.

YÜKSEK BOYUTLU MODEL GÖSTERİLİMİ İLE VERİ BÖLÜNTÜLEME YÖNTEMİNİN KOŞUTLAŞTIRILMASI

ÖZET

Yalnızca uzaydaki belli noktalardaki değerleri verilmiş çok değişkenli bir $f(x_1, x_2, \dots, x_N)$ işlevine alışılacelmiş yöntemlerle içdeğerbiçim işlemi yapılması boyut sayısı arttığında birer başbelası durumuna gelir. Bu tür işlevler için doğrudan bilgisayar programcılığı ile çözüm aramak yerine ilk olarak bu işlevleri bilgisayar programlaması açısından daha kolay ele alınacak, matematiksel olarak, etkili bir yapıya getirmek gerekir. Bu amaçla bu işleve yaklaştırmı yapan bir böl–ve–yönet algoritması geliştirilmiştir. Bu yaklaştırmı sayesinde çok değişkenli f işlevi çok daha düşük boyutlu terimlerle ifade edilebilmektedir. Bu yaklaştırmı Yüksek Boyutlu Model Gösterilimi (YBMG) adı verilmektedir. Bu yöntem çeşitli çalışmalarla başarılı bir şekilde uygulanmıştır. Fakat bu yöntem bu haliyle büyük veri hacmine sahip problemler üzerinde uygulanamaz. Problemdaki boyut sayısı ve boyutlardaki düğüm noktaları sayıları arttığında veri hacmi öyle büyür ki alışılacelmiş PC’ler verinin gereksinim duyduğu yüksek RAM sığasını karşılayamaz. Diğer bir önemli problem de YBMG terimlerini hesaplamakta kullanılan eşitliklerin yapılarıdır. Eşitlikler için yazılmış algoritmadaki döngü sayıları problemdeki boyut sayısına bağımlıdır. Bu çalışmada ilk olarak YBMG terimlerini hesaplamakta kullanılan eşitlikler iyileştirilmiştir. Bu iyileştirme sonucunda eşitliklerin problemdeki boyut sayısına bağımlılığı ortadan kaldırılmıştır. İyileştirilmiş eşitlikler sayesinde yöntem koşutlaştırmaya uygun bir hale getirilmiştir. Son olarak da yöntemin koşutlaştırmasının başarımı çözümlenmiştir.