



التقنية المثالية العامة لمعالجة البيانات الكبيرة

أنس عبدالقادر محمد هادي

رسالة مقدمة لنيل درجة الدكتوراه

علوم الحاسبات

إشراف

أ. د. كمال منصور جمبي

كلية الحاسبات و تقنية المعلومات

جدة - جامعة الملك عبدالعزيز

شعبان ١٤٤٠ - أبريل ٢٠١٩



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



التقنية المثالية العامة لمعالجة البيانات الكبيرة

أنس عبدالقادر محمد هادي

المستخلص

يعد اقتراح وتطوير استراتيجيات تحوير جديدة و ضبط معلمات التحكم لتحسين أداء التحسين للتطور التفاضلي (Differential Evolution) من الدراسات المهمة في الوقت الحاضر. لذلك فإن المساهمات الرئيسية لهذه الأطروحة تتمحور حول الاتجاهات الستة التالية:

الاتجاه الأول هو تقديم استراتيجية تحوير أقل جشعاً (ord_best) مزودة بإمكانية استكشاف معززة. في الاتجاه الثاني تقدم استراتيجية طفرة أكثر جشعاً (ord_pbest) مزودة بإمكانية استغلال محسنة. تعتمد كل من استراتيجيتي التحوير المقترحة على ترتيب ثلاثة متجهات محددة من الجيل الحالي لتغيير المتجه المستهدف، حيث يتم استخدام الاختلافات الموجهة لتقليد سلوك نزول التدرج لتوجيه البحث نحو حلول أفضل. حيث يتم اختيار المتجهات الثلاثة بشكل عشوائي لتعزيز قدرة الاستكشاف للخوارزمية. من ناحية أخرى، تم تصميم استراتيجية (ord_pbest) لتعزيز القدرة على الاستغلال حيث يتم اختيار متجهين بشكل عشوائي ويتم اختيار الثالث من أفضل المتجهات العامة. بناءً على استراتيجيتي التحوير المقترحة تم تقديم خوارزميتين مطورتين من خوارزمية التطور التفاضلي. الاتجاه الثالث تم من خلال تقديم نهج جديد لتكليف المعلمات كنهج بديل للتكليف معلمات التحكم المستخدمة حالياً. يتكون النهج المقترح من إعدادين مختلفين لمعلمي التحكم: عامل القياس (F) ومعدل التقاطع (Cr). وتتمثل فائدة هذا النهج في إثبات أن الخوارزمية شبه التكييفية أفضل من الخوارزمية العشوائية البحتة أو الخوارزمية التكييفية الكاملة والذاتية. الاتجاه الرابع هو إطار تهجين منخفض المستوى يمكن دمج استراتيجيتي التحوير مع استراتيجيات تحوير أخرى لتعزيز قدرات البحث في المسائل الصعبة والمعقدة. يقترح الاتجاه الخامس إطارين للتهجين عالي المستوى بين الخوارزميات المقترحة وخوارزميات التحسين الأخرى. الأول هو إطار تهجين بين الخوارزميات المقترحة وإصدار معدل من استراتيجيات تطور مصفوفة التغير (CMA-ES) حيث تخضع النسخة المعدلة من (CMA-ES) لعملية التقاطع لتحسين قدرة الاستكشاف للإطار المقترح. من ناحية أخرى ستعمل كلتا الخوارزميتين في نفس الوقت على نفس المجموعة، ولكن سيتم تعيين المزيد من الأفراد بشكل تدريجي لخوارزمية الأداء الأفضل. في إطار التهجين الثاني ستكون الخوارزميات المقترحة هي الخوارزميات القائمة على المجموعات، في حين أن النسخة المعدلة من البحث عن مسار متعدد (MTS) ستكون خوارزمية البحث المحلي. الاتجاه السادس هو تقديم مفهوم محسّن للتوزيع وإعادة التجميع لحل مشكلات التحسين العامة واسعة النطاق،



حيث يتم تقسيم الأبعاد بشكل عشوائي إلى مجموعات، ثم يتم حل كل مجموعة على حدة. أخيراً، من أجل التحقق من أداء الخوارزميات المقترحة وتحليلها، أجريت تجارب رقمية باستخدام معايير (CEC2013) و (CEC2017) لتقييم الأداء. بالإضافة إلى معايير (CEC2010) و (CEC2013) المصممة لحل مشكلات التحسين العامة واسعة النطاق.



GLOBAL OPTIMIZATION TECHNIQUE FOR BIG DATA PROCESSING

By

Anas Abdulqader Mohammed Hadi

**A thesis submitted for the requirements of the degree of
Doctor of Philosophy in Computer Science**

Supervised By

Prof. Kamal M. Jambi

Faculty of Computing and Information Technology

King Abdulaziz University - Jeddah

Shaban 1440H – April 2019G

Global Optimization Technique for Big Data Processing

Anas Abdulqader Mohammed Hadi

Abstract

Proposing new mutation strategies and adjusting control parameters to improve the optimization performance of differential evolution (DE) is an important research study. Therefore, the main contribution of this thesis goes in the following six directions: The first direction is introducing a less greedy mutation strategy with enhanced exploration capability, named *DE/current-to-ord_best/1* (*ord* stands for ordered) or *ord_best* for short.

In the second direction, we introduce a more greedy mutation strategy with enhanced exploitation capability, named *DE/current-to-ord_pbest/1* (*ord_pbest* for short). Both of the proposed mutation strategies are based on ordering three selected vectors from the current generation to perturb the target vector, where the directed differences are used to mimic the gradient descent behavior to direct the search toward better solutions. In *ord_best*, the three vectors are selected randomly to enhance the exploration capability of the algorithm. On the other hand, *ord_pbest* is designed to enhance the exploitation capability where two vectors are selected randomly and the third is selected from the global p best vectors. Based on the proposed mutation strategies, *ord_best* and *ord_pbest*, two DE variants are introduced as EDE and EBDE, respectively.

The third direction is introducing a new semi-parameter adaptation approach (SPA) as an alternative adaptation approach for the selection of control parameters. The proposed approach consists of two different settings for the two control parameters: the scaling factor (F) and the crossover rate (Cr). The benefit of this approach is to prove

that the semi-adaptive algorithm is better than the pure random algorithm or the fully adaptive or self-adaptive algorithm.

The fourth direction is a low-level hybridization framework. The proposed mutations can be combined with other mutations to enhance their search capabilities on difficult and complicated optimization problems.

The fifth direction is proposing two high-level hybridization frameworks between the proposed algorithms and other optimization algorithms. The first is a hybridization framework between the proposed algorithms and a modified version of the Covariance Matrix Adaptation Evolution Strategy (CMA-ES). The modified version of CMA-ES undergoes the crossover operation to improve the exploration capability of the proposed framework. Both algorithms will work simultaneously on the same population, but more populations will be assigned gradually to the better performing algorithm. The second hybridization framework is a new memetic hybridization framework, where the proposed algorithms will be the population-based algorithms, while a modified version of Multiple Trajectory Search (MTS) will be the local search algorithm.

The sixth direction is introducing an improved divide and conquer concept for solving large-scale global optimization problems, where the dimensions are randomly divided into groups, and each group is solved separately.

Finally, in order to verify and analyze the performance of this work, numerical experiments were conducted using CEC2013 and CEC2017 benchmarks. The performance was also evaluated using CEC2010 and CEC2013 benchmarks designed for Large Scale Global Optimization.