جامعة الموصل
كلية علوم الحاسوب والرياضيات

طريقة اختيار الجينات عالية الأبعاد بالاعتماد على تحسين خوارزمية الخفافيش الثنائية

فراس أحمد يونس الطائي

رسالة ماجستير

بإشراف
الأستاذ المساعد الدكتور
عمر صابر قاسم

1441هـ 2020م

المستخلص

تضم العديد من التطبيقات ولا سيما التطبيقات الطبية المتعلقة بدراسة الجينات كماً هائلا من البيانات المتعلقة بوصف هذه الجينات، وأن الحاجة إلى نتائج هذه البيانات دفعت بالكثير من العلماء إلى البحث عن الطرق التي تعمل على استخلاص تلك المعلومات من خلال اختيار أهم الميزات الموجودة في تلك البيانات، بالاعتماد على مفهوم اختيار الميزة Feature Selection، الذي يعد أحد أهم الأساليب المستخدمة في تحسين عملية تصنيف البيانات الكبيرة في العقود الثلاث الماضية، وبالاعتماد على خوارزميات الاسراب Swarm Algorithms المستوحاة من سلوك بعض الكائنات الحية، مثل خوارزمية الخفاش Bat Algorithm (BA) بعد تحويلها من الفضاء المستمر Continuous Space إلى الفضاء المتقطع Discrete Space، من خلال مفهوم Binary BA الذي يتعامل مع المتجه الثنائي الذي قيمه اما 0 او 1.

تم في هذه الدراسة اقتراح طريقة لتصنيف البيانات تتكون من ثلاثة مراحل أساسية: إذ تتمثل المرحلة الأولى باستخدام تقنية المعلومات المتبادلة (MI) التي يتم من خلالها ترتيب الميزات حسب أهميتها، أما المرحلة الثانية فيتم من خلالها استخدام إحدى الخرائط الفوضوية Chaotic Map والتي تعمل على تحديد القيمة الابتدائية للمعلمات المستخدمة في خوارزمية BBA، بدل القيم العشوائية التي قد تسبب فوضى بعد عدد من التكرارات اثناء تنفيذ الخوارزمية، أما المرحلة الثالثة فيتم من خلالها استخدام مصنف Naive Bayes لتصنيف البيانات الناتجة من المرحلة الثانية. وقد أثبتت نتائج الطريقة المستخدمة (المهجنة) CBBA الكفاءة والدقة العالية مقارنة بالخوارزمية الاعتيادية BBA بعد تطبيقها على مجموعتين من البيانات هما ال prostate،Ovarian.

UNIVERSITY OF MOSUL
COLLEGE OF COMPUTER SCIENCES
AND MATHEMATICS

Method for selecting high-dimensional genes based on the improvement of the Binary Bat Algorithm

Firas Ahmed Younis Al-Taie

M.Sc./Thesis

Mathematics/Computational

Supervised by

Assistant Professor

Dr. Omar Saber Qasim

2020 A.D. **1441 A.H.**

Abstract

Many applications include especially medical applications related to the study of genes a big amount of dataset related to the description of these genes, and the need of results of these bigdata has driven many scientists to search by that methods to extract that information by choosing the most important features found in that bigdata by depending on the concept of feature selection, which is one of the most important techniques used to improve the classifying big data in the past three decades, and by depending on the evolutionary algorithms that are inspired by the behavior of some organisms, like the bat algorithm (BA) after being transform it from continuous space to discrete space through the binary BA concept which dealing with binary vector {0,1}. In this study, a method for classifying big data consisting of three basic stages has been proposed: The first stage is represented by the use of mutual information technique (MI) through which the arrange the features according to their effect on the classification process, while the second stage, it is applied through the use of chaotic map which works to determine the initial value of the parameters used in the BBA algorithm instead of random values that may cause chaos after a number of iterations during the applying of the algorithm. As for the third stage, a Naive Bayes classifier is used to classify the subset data which results from the second stage. The results of the proposed CBBA method demonstrated efficiency and high accuracy compared to the normal BBA algorithm.